Defense Nuclear Agency
Alexandria, VA 22310-3398

AD-A286 891

DNA-TR-93-173-V3

The Effect of MOPP4 on M198 Howitzer Crew Performance

Volume 3—Validation of Subject Matter Expert (SME) Performance Estimation Method

J. Thomas Roth
Micro Analysis and Design
4900 Pearl East Cir # 201E
Boulder, CO 80301-6108

Gene E. McCellan
Pacific-Sierra Research Corp.
Washington Operations
1400 Key Blvd
Suite 700
Arlington, VA 22209

May 1996

Technical Report

CONTRACT Nos. DNA 001-90-C-00139
and DNA 001-90-C-00164

Approved for public release;
distribution is unlimited.

96-01708

Destroy this report when it is no longer needed. Do not return to sender.

PLEASE NOTIFY THE DEFENSE NUCLEAR AGENCY, ATTN: CSTI, 6801 TELEGRAPH ROAD, ALEXANDRIA, VA 22310-3398, IF YOUR ADDRESS IS INCORRECT, IF YOU WISH IT DELETED FROM THE DISTRIBUTION LIST, OR IF THE ADDRESSEE IS NO LONGER EMPLOYED BY YOUR ORGANIZATION.

DISTRIBUTION LIST UPDATE

This mailer is provided to enable DNA to maintain current distribution lists for reports. (We would appreciate your providing the requested information.)

- ☐ Add the individual listed to your distribution list.
- ☐ Delete the cited organization/individual.
- ☐ Change of address.

NOTE:
Please return the mailing label from the document so that any additions, changes, corrections or deletions can be made easily. For distribution cancellation or more information call DNA/IMAS (703) 325-1036.

NAME: ______________________________

ORGANIZATION: ______________________________

OLD ADDRESS	**CURRENT ADDRESS**

TELEPHONE NUMBER: (   ) ______________________________

DNA PUBLICATION NUMBER/TITLE	**CHANGES/DELETIONS/ADDITIONS, etc.)** *(Attach Sheet if more Space is Required)*

DNA OR OTHER GOVERNMENT CONTRACT NUMBER: ______________________________

CERTIFICATION OF NEED-TO-KNOW BY GOVERNMENT SPONSOR (if other than DNA):

SPONSORING ORGANIZATION: ______________________________

CONTRACTING OFFICER OR REPRESENTATIVE: ______________________________

SIGNATURE: ______________________________

CUT HERE AND RETURN

DEFENSE NUCLEAR AGENCY
ATTN: IMAS
6801 TELEGRAPH ROAD
ALEXANDRIA, VA 22310-3398

DEFENSE NUCLEAR AGENCY
ATTN: IMAS
6801 TELEGRAPH ROAD
ALEXANDRIA, VA 22310-3398

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	960501	Technical 901001 – 940630

4. TITLE AND SUBTITLE
The Effect of MOPP4 on M198 Howitzer Crew Performance
Volume 3—Validation of Subject Matter Expert (SME) Performance Estimation Method

5. FUNDING NUMBERS
C - DNA 001-90-C-0139
DNA 001-90-C-0164
PE - 62715H
PR - RM
TA - RH
WU - DH303390
DH319720

6. AUTHOR(S)
J. Thomas Roth (MA&D), Gene E. McClellan (PSR) and Arthur P. Deverill (ARES)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Micro Analysis and Design
4900 Pearl East Cir # 201E
Boulder, CO 80301-6108

Pacific-Sierra Research Corp.
Washington Operations
1400 Key Blvd, Suite 700
Arlington, VA 22209

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Defense Nuclear Agency
6801 Telegraph Road
Alexandria, VA 22310-3398
RAEM/Kehlet

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
DNA-TR-93-173-V3

11. SUPPLEMENTARY NOTES

This work was sponsored by the Defense Nuclear Agency under RDT&E RMC Code B4662D RM RH 00124 RARP 3400A 25904D.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

Validation of a method for obtaining estimates of task performance change under stressor conditions is extended by comparing Subject Matter Expert (SME)-derived data with measured data on the performance of M198 howitzer crews during live-firing. Fair agreement between the two bodies of data is found, given the limitations of the data available. The SME-estimate method for obtaining performance change data is conditionally valid. Suggestions for improving the estimation methodology are given.

14. SUBJECT TERMS

MOPP4	Subject Matter Expert
Combat Models	M198 Howitzer Crews
Human Performance	Performance Estimation

15. NUMBER OF PAGES 68

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

NSN 7540-280-5500

Standard Form 298 (Rev.2-89)
Prescribed by ANSI Sts. 239-18
298-102

CLASSIFIED BY:

N/A since Unclassified.

DECLASSIFY ON:

N/A since Unclassified.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) (Continued)

ARES Corp
1800 North Kent Street
Suite 1230
Arlington, VA 22209

SUMMARY

This report contrasts the measured performance of selected M198 howitzer crew tasks in the full MOPP 4 personal protective ensemble with Subject Matter Experts' (SMEs') estimates of performance on analogous tasks after various times in MOPP 4, obtained using a structured questionnaire methodology. The analyses reported here take advantage of a rare opportunity to gather measured data on changes in task performance by military personnel under real-world stressor conditions. Collection of the actual performance-change data was the primary purpose of the work reported here and in other volumes of this report (McClellan, Deverill, and Matheson, 1994 and McClellan, Matheson, and Deverill, 1994). Use of the SME-estimate method was a secondary, but important, goal during data collection. The analyses reported in this volume were performed to extend the validation of the structured questionnaire method for obtaining SME estimates of human-task performance degradation in the presence of battlefield stressors.

Two methods of comparing measured and SME-estimate data are used. In the first method, performance metrics derived from measured task performance data for 11 tasks are statistically compared with SME-derived performance metrics for similar tasks at 1, 2, and 4 hours in MOPP 4. The second method statistically compares the slope terms of regression equations computed from measured data, describing performance as a function of time in MOPP 4 for the 11 tasks, with the slope terms of similar equations derived from the SME-estimate data.

SMEs' estimates of performance change due to enclosure in the MOPP 4 ensemble are generally similar to performance change data derived from observation. However, in point-by-point comparison, interesting differences arise. In about half the cases where comparisons were made, SME performance change estimates differed statistically from observationally-derived data. SMEs overestimate performance degradation (relative to measured data) more frequently than they underestimate. For one task, SMEs underestimate performance decrements for all three periods in MOPP 4, the underestimate increasing in magnitude with longer time in MOPP 4. For several other tasks, SMEs tend to consistently overestimate performance decrements due to MOPP 4 conditions, across the three estimation times. There was some tendency for the amount of the overestimation of performance decrements to increase for longer times in MOPP 4 for these tasks, but this was not consistent. SMEs' tendency for under- *versus* over-estimation of performance decrements was related to the rated demand for physical ability of tasks. SMEs appear to overestimate performance degradation due to wearing MOPP 4 for tasks that inherently demand larger amounts of physical ability and underestimate performance decrements for less physically-demanding tasks.

Comparison of the slope terms of the regression equations based on measured data and those derived from SME-estimate data for the 11 tasks indicated agreement for 7 tasks and statistically significant differences for 4 tasks. For each of the tasks showing differences, and for one additional task, examination of the data indicated that measured task performance was increasing as a function of increased time in MOPP 4, in contrast to expectations about the performance-degrading effects of MOPP 4 on performance. Each of these five tasks is performed by either the Gunner or the Assistant Gunner of the howitzer crew. For the remaining six tasks, expected performance decrements as a function of time in MOPP 4 were found in the measured

data, and the slope terms of the regression equations based on measured data did not differ statistically from those derived from SME-estimate data. All SME estimates of performance showed monotonic decreases in performance as a function of increased time in MOPP 4. The observed differences between measured and SME-estimate data may be due to the relative inexperience of crewmembers who were performing the tasks done by the Gunner and Assistant Gunner crew positions. The noted increases in performance for these tasks may indicate that crewmembers were still learning the component skills required to carry out the tasks in MOPP 4. The effects of this continued learning on performance may have overshadowed the performance-degrading effects of operating in MOPP 4.

Based on these findings, it is concluded that the SME-estimate method for obtaining data on performance change as a function of stressor exposure is conditionally validated. While SMEs had a tendency to overestimate the effect on performance of enclosure in the MOPP 4 ensemble for physically demanding tasks and to underestimate the effect for physically undemanding tasks, in general the SMEs made predictions of performance change that are more or less accurate when compared to observationally-measured howitzer-crew task performance. These results were found despite the less-than-ideal characteristics of the SMEs who gave performance estimates, and some limitations in the observationally-derived task performance data. The findings support the continued use of the SME-estimate method described for obtaining data on which to base predictions of personnel performance change due to exposure to battlefield stressors.

Preface

This report was prepared for the Radiation Risk/Safety Program at the U.S. Defense Nuclear Agency (DNA). DNA's technical monitors for this project were Mr. Robert A. Kehlet and Dr. Robert W. Young of the Environments and Modeling Division. Data collection and analysis were supported jointly by Micro Analysis and Design, Inc. (MA&D) and the ARES Corporation with funding from DNA through contracts DNA001-90-C-0139 and DNA001-90-C-0164, respectively. Pacific-Sierra Research Corporation participated with funding from these same contracts through MA&D subcontract SC 102 and ARES subcontract ARES-PSR-90-C-001.

Data were collected by DNA researchers from ARES, EAI Corporation, and PSR on a non-interference basis during a four-week exercise conducted in August of 1992 by the U.S. Army Human Engineering Laboratory (HEL) at Aberdeen Proving Ground, Maryland. The exercise, *Assessment of Towed Artillery (M198) Crew Performance in NBC Protective Clothing*, was directed by Mr. Orest Zubal of HEL, whose cooperation is gratefully acknowledged. The exercise was funded by the *Psychological and Physiological Effects of the NBC Environment and Sustained Operations on Systems in Combat* (P^2NBC^2) Project Office of the U.S. Army. Mr. Kehlet and Mr. Don Cunningham, P^2NBC^2 Program Manager, U.S. Army Chemical School, were instrumental in arranging on short notice for the DNA team to take advantage of this important data collection opportunity. The analyses reported in this volume were performed by MA&D researchers, with assistance from PSR and ARES personnel.

Volume I of this report provides data on the effects of MOPP 4 on the performance of howitzer emplacement and displacement activities and on rates of fire. Volume II provides detailed analyses of MOPP 4-induced degradation of the performance of individual crewmember tasks during fire missions.

CONVERSION TABLE

Conversion factors for U.S. customary to metric (SI) units of measurement

MULTIPLY *THIS* UNIT	——▶ *BY* ——▶	TO GET *THIS* UNIT
TO GET *THIS* UNIT	◀—— *BY* ◀——	DIVIDE *THIS* UNIT
angstrom	1.000 000 X E -10	meter (m)
atmosphere (normal)	1.013 250 X E +2	kilo pascal (kPa)
bar	1.000 000 X E +2	kilo pascal (kPa)
barn	1.000 000 X E -28	meter^2 (m^2)
British Thermal Unit (thermochemical)	1.054 350 X E +3	joule (J)
calorie (thermochemical)	4.184 000	joule (J)
cal (thermochemical)/cm^2	4.184 000 X E -2	megajoule/m^2 (MJ/m^2)
curie	3.700 000 X E +1	*giga becquerel (GBq)
degree (angle)	1.745 329 X E -2	radian (rad)
degree Fahrenheit	$t_k = (t°f + 459.67) / 1.8$	degree kelvin (K)
electron volt	1.602 190 X E -19	joule (J)
erg	1.000 000 X E -7	joule (J)
erg/second	1.000 000 X E -7	watt (W)
foot	3.048 000 X E -1	meter (m)
foot-pound-force	1.355 818	joule (J)
gallon (U.S. liquid)	3.785 412 X E -3	meter^3 (m^3)
inch	2.540 000 X E -2	meter (m)
jerk	1.000 000 X E +9	joule (J)
joule/kilogram (j/kg) (radiation dose absorbed)	1.000 000	Gray (Gy)
kilotons	4.183	terajoules
kip (1000 lbf)	4.488 222 X E +3	newton (N)
kip/in^2 (ksi)	6.894 757 X E +3	kilo pascal (kPa)
ktap	1.000 000 X E +2	newton-second/m^2 (N-s/m^2)
micron	1.000 000 X E -6	meter (m)
mil	2.540 000 X E -5	meter (m)
mile (international)	1.609 344 X E +3	meter (m)
ounce	2.834 952 X E -2	kilogram (kg)
pound-force (lbs avoirdupois)	4.448 222	newton (N)
pound-force inch	1.129 848 X E -1	newton-meter (N•m)
pound-force	1.751 268 X E +2	newton-meter (N•m)
pound-force/ft^2	4.788 026 X E -2	kilo pascal (kPa)
pound-force/inch^2 (psi)	6 894 757	kilo pascal (kPa)
pound-mass (lbm avoirdupois)	4.535 924 X E -1	kilogram (kg)
pound-mass-foot^2 (moment of inertia)	4.214 001 X E -2	kilogram-meter^2 (kg•m^2)
pound-mass/foot^3	1.601 846 X E +1	kilogram-meter^3 (kg•m^3)
rad (radiation dose absorbed)	1.000 000 X E -2	Gray (Gy)
roentgen	2.579 760 X E -4	coulomb/kilogram (C/kg)
shake	1.000 000 X E -8	second (s)
slug	1.459 390 X E +1	kilogram (kg)
torr (mm Hg, 0° C)	1.333 220 X E -1	kilo pascal (kPa)

* The becquerel (Bq) is the SI unit of radioactivity; 1 Bq = 1 event/s.

Table of Contents

Appendix		Page

FIGURES

TABLES

SECTION 1
INTRODUCTION

One tool frequently used to plan and prepare for military operations is *combat modeling*: using abstract, dynamic simulations of battle (and, sometimes, associated logistics play) under selected conditions, given assumptions about forces, terrain, and other factors. While combat models are a valuable adjunct to operational planning efforts, they have often omitted one of the more critical variables in combat: human performance. In many, if not most, combat models, human performance is treated as a constant.

It is well-recognized that the absence of human performance variability in combat models is a critical shortcoming. To date, however, it has been difficult to incorporate this critical source of variability into such models, primarily due to the lack of data to accurately depict human performance changes under particular stressor conditions. The work described here is part of a series of studies, sponsored by the Defense Nuclear Agency (DNA), whose goal is to provide the capability to represent human performance variability in combat models.

One method of obtaining data on human performance variability, especially useful when it is difficult to measure performance under actual situations representing battlefield stressors, is to have Subject Matter Experts (SMEs) make estimates of performance change when specified levels of stressors are in effect. The analyses reported here take advantage of a rare opportunity to gather measured data on changes in task performance by military personnel under real-world stressor conditions. Collection of the actual performance-change data was the primary purpose of the work reported here and in other volumes of this report (McClellan, Deverill, and Matheson, 1994 and McClellan, Matheson, and Deverill, 1994). Use of the SME-estimate method was a secondary, but important, goal during data collection. The analyses reported in this volume were performed to extend the validation of the structured questionnaire method for obtaining SME estimates of human-task performance degradation in the presence of battlefield stressors.

1.1 METHODOLOGY OVERVIEW.

The absence of human performance variability in combat models is not due to a lack of appreciation of the importance of this variable. Rather, as remarked above, it is due to a lack of valid and reliable data for representing changes in human performance under various stressor conditions. Such data have not generally been available for a number of reasons. First, it is unethical (if not illegal) to apply extreme levels of certain stressors (notably nuclear, chemical, and biological conditions or agents) to human subjects in order to study their effects. While some generalizations of such effects to humans from animal models can be made, these are limited and not always valid. In some cases, such extrapolations are not possible. For example, it is impossible to extrapolate stressor effects on human cognitive processes (and associated tasks) from animal models, since human cognitive processes are both qualitatively and quantitatively much more elaborate than those of infrahuman animals.

Second, even where ethical considerations do not apply (other than the obligation to safeguard human subjects from undue stress), obtaining data on humans' performance-related responses to stressors is difficult and costly. Both baseline (unstressed) and experimental observations of task performance must be made to characterize performance changes; often, numerous levels of stressor effects must be engendered in order to fully assess the effects of a stressor on human task performance. Also, the performance-degrading effects of stressors may depend on task type, requiring that many different tasks be included in experimental investigations. Such human experimentation, using real-world tasks under realistic conditions, is resource-intensive and costly to perform.

Finally, there is the issue of extrapolating stressors' effects on performance from one human task to another. Given that actual, measured data on the effects of a stressor on performance are available for some tasks, it is not always straightforward to relate such effects to other, dissimilar tasks. This issue is being pursued, with some success, in a parallel effort of the DNA Radiation Risk/Safety Program, using an ability-demand taxonomic approach for characterizing human tasks and generalizing stressor effects across tasks. For some preliminary findings on this aspect of the research, see Roth (1991, 1992).

Alternate methods have been devised for assessing stressor effects on human task performance. In general, these methods involve Subject Matter Expert (SME) assessments of the effects that signs and symptoms of stressors' effects will have on task performance (see Anno, Wilson, and Dore, 1984 for a discussion of early applications of this method using the physiological effects of nuclear radiation exposure as a stressor).

One goal of these studies is to use data derived from the SME-estimation method, and generalized using the task ability-demand taxonomy, to prepare data suitable for use in combat models to represent changes in human task performance in response to battlefield stressors. Attaining this goal will enable an increased degree of realism in the play of combat models and, thus, potentially more accurate assessments of the outcomes of various alternatives considered in planning and preparing for combat

In typical uses of the SME-estimation approach, a dose-response continuum is defined based on various levels of exposure to a stressor. The dose-response continuum is frequently multidimensional, as in the case of the effects of nuclear radiation exposure, where six physiological response dimensions have been defined (Anno, Wilson, and Baum, 1985). The dose-response continuum need not necessarily be multidimensional, however. In the present work, the effects of a stressor (wearing the full chemical-protective ensemble [MOPP 4] for extended periods of time) are characterized using a single dimension: time in MOPP 4. While there are certainly multiple physiological dimensions of effects of this stressor on performance, each can be represented in a surrogate sense by the amount of time in MOPP 4 under specified environmental conditions. Thus, it is convenient to use time in MOPP 4 as a common surrogate for all of the effects.

Severity scales are defined for each dimension, or symptom category, of a dose-response continuum, generally ranging in five levels from no effect at all to the worst possible physiological effect for the symptom category. The multidimensional specification of the severity level for each

symptom category caused by a stressor at a given dose and time after exposure is called a symptom complex. SMEs are asked to estimate the performance capability of a typical crewmember who is suffering the signs and symptoms of illness described by such a symptom complex.

To avoid inordinate requirements of time and sustained effort from the SMEs, only a limited number of symptom complexes can be used in each performance assessment. For example, the symptom complex for nuclear radiation exposure includes six symptom categories each with five defined severity levels for a total of $5^6 = 15{,}625$ unique symptom complexes. It is unlikely that any SME would have the patience—or be afforded the time—to make so many judgments. Therefore, a set of symptom complexes are selected to best represent the dominant dose and time dependence of stressor response and to reasonably span the multidimensional space of possible complexes. Generally, up to 30 or 40 discrete symptom complexes are used to obtain SME estimates of stressor effects on performance (Anno, Wilson and Dore, 1984; Anno, Dore, Roth, LaVine, and Deverill, 1994).

Questionnaires are prepared for SMEs' response in terms of the effects of the particular signs and symptoms of illness (or other stressor effect), including descriptions of the tasks for which estimates are to be made and the signs-and-symptoms conditions under which performance is to be estimated. A discussion of the details of the SME-estimate method used in the present assessment is included in the second section of this report.

The SME-estimation method has been used on several occasions to obtain data on stressor effects on performance. Anno, Wilson, and Dore (1984), in the DNA Intermediate Dose Program (IDP), obtained estimates of the effects of exposure to nuclear radiation for tasks performed by several types of Army teams or small units. More recently, Anno, Dore, Roth, LaVine, and Deverill (1994) used similar methods to obtain estimates of the effects of two chemical agents (GB and HD), nuclear radiation, and time in MOPP 4 conditions on the performance of Army Field Artillery Firing Section tasks, and estimates of the effects of GB and HD exposure and time in MOPP 4 on the performance of (light) Infantry tasks.

While the SME-estimate method for obtaining estimates of stressor effects on performance is conceptually compelling, a comprehensive direct validation of these estimates has proven elusive, for the same reasons cited above regarding the paucity of data on actual stressor effects on human performance. One effort that demonstrated some convergent validity for the SME-estimate method is that of McClellan and Wiker (1985). In this effort, a (partial) mapping between the sign-and-symptom complexes related to nuclear radiation exposure (developed in the IDP effort) and symptoms of seasickness (resulting from empirical observation) was accomplished. (Many of the symptoms of acute seasickness resemble the prodromal symptoms of exposure to moderate doses of nuclear radiation.) Empirical data on task performance decrements resulting from seasickness were compared to SME estimates of performance decrements on globally similar tasks from the IDP research. McClellan and Wiker (1985) concluded that "...estimates by Army operational personnel of performance decrements from radiation sickness are quite similar in magnitude to the measured [performance] decrements of Coast Guardsmen during motion sickness." Thus, some amount of (convergent) validity for the SME-estimate method was demonstrated by these investigators.

1.2 THE PRESENT EFFORT.

The Army exercise which this report concerns provided an opportunity to further, and more directly, validate the SME-estimate method for obtaining data on task performance decrements resulting from exposure to stressors. Before discussing details of the validation effort, a general outline of the context and procedures of this work is appropriate.

The data on which analyses in this report are based were gathered by a DNA contractor team during a 4-week exercise in August of 1992. Data were collected at U.S. Army's Aberdeen Proving Ground (APG), Maryland, during a Field Artillery live-firing exercise under the auspices of a program entitled *Psychological and Physiological Effects of the NBC Environment and Sustained Operations on Systems in Combat*, abbreviated P^2NBC^2. Three Marine Corps M198 (155 mm) towed howitzer crews participated in the exercise. Each participating crew was an intact team consisting of ten members: the Chief of Section, Gunner, Assistant Gunner, Radio-telephone Operator (RTO), and six Cannoneers.

Each crew participated in both Battle-Dress Uniform (BDU, also referred to here as Baseline) conditions, and in MOPP 4, the full chemical-protective ensemble. Each crew spent one week in preparation and training for the exercise and conducted a series of fire missions on three days of the week following the preparation period (one crew was preparing during the week another crew participated in live-firing).

Planned live-fire exercises for each firing day included 17 fire missions wherein a total of 89 inert rounds were to be fired by a crew. Nominally, crews first emplaced their howitzers and performed seven fire missions (firing a total of 30 rounds; two of these fire missions were "high angle" missions, using higher elevations of the cannon tube, the remaining five were "normal" missions using elevation angles of less than 1 radian). Then, the crew performed a road march for resupply. Following the road march, crews were to re-emplace the weapons and fire an additional 10 fire missions. Of these latter 10 missions, nine were 3-to-5 round "normal" missions, and one was a "zone and sweep" fire mission with 25 rounds fired on a 5 x 5 grid of aim points.

Each crew performed for one day in BDU conditions (no chemical-protective equipment worn), and at least one day in the full MOPP 4 ensemble. Two of the crews performed an additional day in MOPP 4, using an experimental crew-position rotation scheme that was intended to reduce heat stress on crewmembers (data from these "rotation" observations are not dealt with in this report). The participants' core temperature and heart rate were monitored, and participants were removed from participation when certain physiological limits were exceeded (see McClellan, Deverill, and Matheson, [1994] for details). When a crew was reduced to six members through the removal process, that crew's participation was terminated for the day. Therefore, not all of the planned fire missions were fired by every crew on every day of their participation.

Measured performance data collection during the live-fire exercises was accomplished by three observers, who recorded data on notebook computers. Each of the observers recorded the occurrence of specific, observable events by pressing coded keys on a computer keyboard. Event timing was performed by the computers' internal clocks. Each observer was responsible for observing and recording a specific set of events; some key events were recorded by all of the

recorders to enable later reconciliation of time data across observers. Details of the data collection, reduction, and cross-observer reconciliation processes are found in Volumes I and II of this report.

Crewmembers participating in the exercise also made estimates of crew task performance under MOPP 4 conditions, using the questionnaire-based SME-estimation method mentioned earlier and described in the next section of this report. Each crewmember made two different estimates for each of nine tasks: one estimate was made before the live-firing in MOPP 4 portion of the exercise ("pre" estimate), and one was made following live-firing in MOPP 4 ("post" estimate). The comparison of these SME-estimate data with measured data taken during the live-firing portion of the exercise is discussed in the remainder of this report.

1.3 SOME CAVEATS.

Two cautions must be raised regarding the data, analyses, and conclusions contained in the following sections of this report. The first concerns subject matter expertise. The second is related to the statistical analyses reported here.

It is the usual practice in gathering SME-estimate data of the kind we discuss here to use "true" SMEs to provide performance estimates. For military tasks of the sorts in which we are interested, a "true" SME is usually chosen according to the following criteria:

1. An SME should be of a senior enlisted grade (at least E-6, or Staff Sergeant in Army or Marine Corps grade title) and hold the primary Military Occupational Specialty that usually performs the task(s) for which performance estimates are to be made.

2. An SME should have recent experience (within the last 6 months) as the principal leader of the type of unit that usually performs the task(s) for which performance estimates are to be made.

3. An SME should have had recent experience performing the tasks for which estimates are to be given *both under normal conditions and under conditions similar to those described to induce degraded performance*.

4. An SME should have sufficient literacy skills to be able read and comprehend the performance estimation questionnaires to be used and the written instructions for performing the estimation task.

Normally, a panel of five or more "true" SMEs is convened to perform the estimation task (more is better), but each makes his or her estimates independently. Ideally, before the panel is disbanded, an opportunity is available to question the rationale for SMEs' estimates that are significantly different from the norm for the panel and afford the SME(s) that provided the discrepant estimates an opportunity to reconsider and possibly revise their estimates.

In this effort, the members of the three howitzer crews provided performance-estimate data using the questionnaire assessment process described. Of the 30 members of these crews, none met all three of the criteria above for "true" SMEs. The highest grade represented was E-5 (Sergeant), which was held by only two of the three Chiefs of Section. The Chiefs of Section *did* have recent leadership experience with their firing sections, but other crewmembers did not. The crewmembers in general were relatively inexperienced, but had at some time performed the tasks for which performance estimates were desired. Most of the crewmembers had spent some amount of time in the MOPP 4 ensemble, but data on how recently this had occurred and whether crewmembers had actually performed the specific tasks for which estimates were desired while in MOPP 4 were not recorded. Literacy of the crewmembers was not assessed formally, but all 30 were able to complete the questionnaire.

The point of this digression is that the performance estimates given by the crewmembers in the subject exercise may not be as reliable as those that would (hypothetically) be given by panels of "true" SMEs making performance estimates for the same tasks under the same described conditions. While this does not invalidate the estimates given by the crewmembers, it detracts somewhat from the strength of conclusions that we are willing to draw from the analyses reported here.

The second caution concerns the sizes of samples represented in the analyses. In some cases, particularly with the measured data, relatively few observations were available to characterize performance for any given task at particular times in MOPP 4. Therefore, the confidence bounds for these samples are broad. Conclusions based on these data may not be as firm as if larger samples, resulting in more accurate sample parameter estimates, were available.

SECTION 2
SME PERFORMANCE ESTIMATE DATA

This section describes how the SME performance estimation data were gathered, analyzed, and assessed for use in the validation work. First, the questionnaire assessment method is described in detail. Following this, procedures for administration of the questionnaire are briefly described. Next, scoring and interpretation of the raw data as consistent performance estimates are discussed. Finally, the specific SME-estimation data gathered in this effort are assessed and evaluated for their contribution to the validation analysis effort.

2.1 THE QUESTIONNAIRE ASSESSMENT TECHNIQUE.

The general technique for obtaining SME estimates of task performance is based on use of a performance assessment questionnaire. Each questionnaire page describes a number of tasks for which performance estimates are to be made. Each page contains descriptions of a sign-symptom complex or other conditions the respondent is to consider as in effect for the task performance estimates *on that particular page*. SMEs make independent estimates of performance for each sign-and-symptom complex or set of conditions. They are neither instructed nor required to compare their estimates across different conditions to attain consistency (consistency is evaluated as part of the data reduction process).

Questionnaire pages provide a usual, or nominal, time for performing each task, which the SME is asked to consider in making performance estimates. These nominal times are generally taken from doctrinal or training descriptions of the tasks for which performance estimates are desired. In cases where this information is not available from approved publications, an independent panel of SMEs is used to estimate nominal times for performing each task, or normal task times are measured under innocuous conditions. A representative questionnaire page, modeled after one used in the SME performance estimation process in this effort, is shown as Figure 2-1. The nine tasks whose titles are shown in Figure 2-1 are the tasks for which performance estimates were made in this effort.

In making their estimates, SMEs are instructed to use a three-step decision process, with reference to the nominal time to perform each task. First they are asked to decide, given the conditions (or signs-and-symptoms) described, whether ypical soldier (or other job performer) can perform the task *at all*. If an SME judges that a task cannot be performed at all under the conditions described, the estimation process for the task in question is complete. In this case, the SME simply marks the rightmost column on the questionnaire page, at the level of the title of the task.

Situation:

M-1

Have been at MOPP4 for 1 hour.
It is daytime.
The outside air temperature is 85 degrees.
The humidity is high.

TASK DESCRIPTIONS M198 (155mm) Howitzer	The usual time for each task (sec)	How long do you think it would take an artilleryman to do each task in this situation?		
		No increase in time	Amount of time (sec)	Can not do it at all
CHIEF OF SECTION Receive fire order and relay fire commands to section	25			
GUNNER Set deflection on sight	4			
Traverse tube, level bubble	7			
Check sight picture, level bubbles	4			
ASSISTANT GUNNER Set QE on range quadrant	4			
Elevate tube, level bubbles	8			
NO. 1 CANNONEER Load projectile and propellant	8			
Close breech and prime firing mechanism	5			
Open breech, inspect bore	5			

How much confidence do you have in your estimates?

None	Not Much	Some	A Lot	Certain

Figure 2-1. Representative SME-estimation questionnaire page.

If the SME judges that a task can be performed given the described conditions, the next step is to decide whether the task can be accomplished in the nominal or typical amount of time, or whether additional time will be needed to accomplish the task, given the conditions described. If an SME decides that the task can be performed in the nominal amount of time, the SME marks the second column on the questionnaire page, across from the task title. If this decision (can do it in the nominal time) is reached, then the estimation process for the task in question is complete.

If an SME judges that a task can be done under the described conditions, but will require additional time, he or she then must make an estimate of the time the task will take to accomplish under the conditions described. The SME now estimates the amount of time it will take to do the task under the conditions described, and writes the number of seconds or minutes estimated for task performance in the third column of the questionnaire page.

No provision is made for estimates of task performance in *less than* the nominal time given. The case where no *additional* time is required for task performance is presumed (tacitly) to include performance requiring less than the nominal time, but this point is left moot in instructions to SMEs on how to perform the estimation task. In some cases, SMEs write-in a time estimate in the third column of a questionnaire page that is less than the given nominal amount of time to perform the task. These anomalies are dealt with on a case-by-case basis, described below.

SMEs are instructed to complete estimates for all tasks on one questionnaire page before moving on to make estimates for tasks performed under different described conditions (on other pages). SMEs are neither explicitly instructed to compare their estimates for performing a task across different conditions (pages) nor told not to do so. Observation of SMEs performing the estimation process reveals that few SMEs cross-check performance estimates from one questionnaire page to another (Anno, Dore, Roth, LaVine, and Deverill, 1994).

At the bottom of each questionnaire page, a "response confidence rating" area is provided (see Figure 2-1). SMEs are requested to circle the boxed phrase that best represents their subjective confidence in the accuracy of the information provided on each page.

In this effort, SMEs gave performance estimates for four described sets of conditions. The conditions shown in Figure 2-1 are typical; the only difference between one questionnaire page and another in this study was the amount of time in the MOPP 4 ensemble that was described. SMEs were asked to make performance estimates for four different conditions of time in MOPP 4: 5 minutes, 1 hour, 2 hours, and 4 hours. In this report, only the data from the estimates for 1, 2, and 4 hours in MOPP 4 are considered. This is because no data were gathered on measured performance of the tasks of interest at less than about 1 hour in MOPP 4. During the exercise, crews spent about the first hour each day emplacing the howitzer, unloading ammunition and other equipment, and preparing to perform fire missions. Typically, the first fire mission on any given day did not begin until about one hour after MOPP 4 was donned (or, for the Baseline condition, after moving the howitzer into the designated firing position and laying the howitzer).

2.2 QUESTIONNAIRE ADMINISTRATION.

In this research, SMEs responded to the performance-estimation questionnaires in a group setting. Each of the three crews was administered the questionnaire separately on two occasions: once before the crew performed live-firing in MOPP 4 and once following the crew's first day of live-firing in MOPP 4. Similar procedures were followed during all questionnaire sessions, with no notable exceptions.

At the beginning of each questionnaire session, the SMEs were given a general orientation to the purpose of the questionnaires and how the data would be used. This motivational material was intended to elicit careful and cooperative responses from the SMEs. Procedures for responding to the questionnaire (including the three-step estimation procedure outlined above) were presented, and questions about the response procedures were answered.

A general scenario was then described, which SMEs were to consider when making their performance estimates. Characteristics of the scenario included:

- The unit is in a combat zone, and has been exposed to an agent that requires assuming a MOPP 4 protective posture.
- The exact nature and persistence of the agent is unknown, so unit members remain in MOPP 4 at least the amount of time for which performance estimates are to be made (on each questionnaire page).
- Soldiers in the unit are well-trained and well-led, and are motivated to do their best.
- It is daytime. The unit is prepared to engage in fire missions (*i.e.*, howitzer is already emplaced and other preparations for conducting fire missions are complete).
- The unit is called upon to perform a fire mission. (No other details of the hypothetical fire mission were given; that is, number of rounds to be fired was not specified, the mission variant [normal, high angle, zone and sweep] was not stated, and other details such as previous activities, round type, fuze, etc. were left tacit.)

SMEs were instructed to consider the performance of a hypothetical "average" artilleryman in making performance estimates. They were specifically told not to think of their own hypothetical performance, nor that of the best or worst soldiers they had ever observed performing the tasks, but to consider the performance of a "typical" artilleryman.

Response procedures and methods for recording responses were then reviewed verbally, and the SMEs were again given the opportunity to pose any questions about the questionnaire or procedures and have the issues clarified. SMEs were instructed to complete one questionnaire page at a time, indicating their confidence in the answers given after performance ratings for all tasks on the page had been made.

SMEs completed the questionnaires at their own pace. Before each SME was dismissed, his questionnaire was examined to make certain that performance estimates for all tasks, at each of the four times in MOPP 4, had been given. This cursory inspection of the questionnaire forms was done to assure that SMEs had not completely ignored response instructions, but was not a detailed screening of the data. Several items of demographic information were also provided by each SME as part of the pre-live fire questionnaire administration process. Detailed demographic characteristics of the members of the three crews are given in Appendix A. A summary of demographic features of the three crews is shown in Table 2-1.

Table 2-1. Selected summary characteristics of crews participating in P^2NBC^2 exercise.

Crew Characteristics	**Crew**		
	1	**2**	**3**
Median Grade of Crewmembers	3	3	3
Highest Grade Among Crewmembers	4	5	5
Median Time in Service of Crewmembers (Years)	1.85	1.05	1.60
Longest Time in Service Among Crewmembers (Years)	3.0	9.5	12.9
Proportion of Crewmembers With Previous Experience in MOPP 4	.9 (N=10)	1.0 (N=10)	.75 (N=8)
Median Number of Times Previously in MOPP 4	1	2	3
Median Longest Estimated Amount of Time in MOPP 4 (Hours)	1.5	2.0	1.5
Mean Number of Key* Crew Positions Previously or Currently Held	1.5	1.3	1.5

* Chief of Section, Gunner, Assistant Gunner, No. 1 Cannoneer

2.3 SCORING AND DATA INTERPRETATION.

The raw performance estimate data provided by the SMEs took one of three forms:

- An indication that a typical artilleryman *could not accomplish a task at all* under the conditions described;
- An indication that a typical artilleryman *could accomplish a task*, and could further accomplish it *in the nominal amount of time* given on the questionnaire, under the conditions described, or

- An indication that a typical artilleryman *could accomplish a task* under the conditions described, *but would require additional time* beyond the nominal task time to accomplish the task. When this response was given, an estimate of the number of seconds that would be required to accomplish the task was also given by SMEs.

In accordance with past procedure for using the questionnaire-based SME estimation method, the raw data were converted into *performance indices* before analysis. The rules for this conversion are:

- An indication that an artilleryman *could not accomplish* the task at all was assigned a score of 0 (no performance).

- An indication that an artilleryman could accomplish the task within the nominal task time was assigned a score of 1 (perfect performance).

- Indications that increased time would be required to perform a task were assigned a performance index score by dividing the nominal time to perform the task (t_0) by the SME-supplied (increased) time to perform the task (t_i). This t_0/t_i ratio ensured that the performance index would fall in the range from 0 to 1, since t_i was always greater than t_0. Longer task times therefore resulted in lower performance index scores.

The data were next tabulated by subject and task (separately for pre- and post-performance questionnaires) and assessed for non-responsiveness and *inversions*. Non-responsiveness was judged to have occurred if an SME consistently gave the same response for all (or almost all) nine tasks for all four times in MOPP 4 (*e.g.*, an SME might simply indicate "cannot do it at all" for all nine tasks over all four conditions of time in MOPP 4, when this is contrary to instructions, expectations, and common-sense reasoning about the effects of MOPP 4 on performance).

Inversions are cases where an SME estimates that task performance for a condition (or sign-and-symptom complex, see Anno, Wilson, and Dore [1984] for further details) is greater than the estimate provided by the same SME for the same task under a condition that should (according to dose-response mappings) result in better performance. In this effort, we looked for cases where an SME's performance estimate for a task for longer amounts of time in MOPP 4 exceeded the same SME's performance estimate for the same task for shorter amounts of time in MOPP 4.

No SME's data were altogether rejected from this analysis for non-responsiveness. In a number of cases, either poor handwriting on the questionnaires, or transcription errors that could not be validly corrected, led to uninterpretable responses. This situation was in no case consistent from one task or questionnaire page to the next for any subject. These cases were treated as missing data for the remainder of the analyses.

Only two cases of apparent inversions (both associated with data of questionable interpretability) were found in the data. Each of these involved a single data point where the performance estimate for 4 hours in MOPP 4 appeared to be higher than that for 2 hours in

MOPP 4. Since these data points were derived from suspect data, they were simply treated as missing data for the remainder of the analyses.

Descriptive statistics were computed and summary data plots were prepared for the data supplied by the SMEs before and after the live-firing exercise, for each of the nine tasks separately, and by crew. The statistics were further segregated by the amount of time in MOPP 4 described on the questionnaire pages. The statistics and plots were used in assessing differences between pre- and post-exercise questionnaire data and between crews. This assessment is described in the following subsection.

2.4 ASSESSMENT OF ADMINISTRATION AND CREW DIFFERENCES.

To determine whether the total SME-estimation data set for each task should be used in the validation analyses, the descriptive statistics and plots just mentioned were used. These were supplemented by analyses that assessed the statistical (as opposed to practical) significance of differences between data provided by SMEs from the three crews on the two different occasions (pre-and post-exercise) where estimates were made. The analyses were conducted separately for each of the nine crewmember tasks for which estimates were made. In all cases, the SME performance estimates for 5 minutes in MOPP 4 were excluded from these and subsequent analyses, since no comparable data points for measured actual performance in MOPP 4 are available for any of the tasks.

Potential differences between crews for the two administrations of the questionnaire were assessed by computing Student's t statistics (Dixon and Massey, 1957) between pre- and post-exercise data sets for each of the nine questionnaire tasks, three times in MOPP 4, and three crews separately. This resulted in 81 t-statistics. Nine of the 81 statistics, or 11 percent, reached significance at the 95 percent level of confidence.

For Crew 1, only one significant difference was found between pre- and post-exercise estimates of performance. This difference was found for the task "Receive and Relay Fire Order," performed by the Chief of Section, at 1 hour in MOPP 4. For this task and crew, the mean "pre-" performance index was .99, the mean "post-" performance index was .92. For Crew 2, significant pre-post differences were found for the following tasks and times in MOPP 4:

- "Traverse Tube," performed by the Gunner, at 1 hour in MOPP 4 (mean "pre-" performance index = .50, mean "post-" performance index = .71);
- "Elevate Tube," performed by the Assistant Gunner, at 4 hours in MOPP 4 (mean "pre-" performance index = .31, mean "post-" performance index = .51); and
- "Close Breech and Prime," performed by the No. 1 Cannoneer, at both 1 and 2 hours in MOPP 4 (at 1 hour, mean "pre-" performance index = .63, mean "post-" performance index = .87; at 2 hours, mean "pre-" performance index = .51, mean "post-" performance index = 84).

For Crew 3, significant pre-post differences were found for these tasks and times in MOPP 4:

- "Traverse Tube," performed by the Gunner, at 1 hour in MOPP 4 (mean "pre-" performance index = .73, mean "post-" performance index = .88); and

- "Elevate Tube," performed by the Assistant Gunner, at 1, 2, and 4 hours in MOPP 4 (at 1 hour, mean "pre-" performance index = .65, mean "post-" performance index = .84; at 2 hours, mean "pre-" performance index = .54, mean "post-" performance index = .70; at 4 hours, mean "pre-" performance index = .44; mean "post-" performance index = 57).

Based on these results, it was considered reasonable to conclude that no consistent or programmatic differences exist between pre- and post-exercise estimates of performance by SMEs. On this basis, we tentatively decided that pooling the pre- and post-exercise estimates for computing regression equations was reasonable. This decision was tentative subject to an examination of the differences in performance estimates between the three crews.

To examine these differences, nine one-way analyses of variance (one for each of the nine tasks) were computed, using crew membership as the classification variable. The analyses were computed on pooled pre- and post-exercise data, since the tentative decision to pool these data had been reached at this point.

Six of the nine analyses of variance reached statistical significance at the 95 percent level of confidence; thus, statistically reliable differences were found among the crews for two-thirds of the tasks for which SME-estimate data were provided. *Post hoc* multiple range tests (Tukey's HSD procedure; Kirk [1968]) were performed to identify which crews' data differed from that of other crews. The experimentwise error rate for each of these tests was set at α=.05.

For the task "Receive and Relay Fire Order," performed by the Chief of Section, Crew 2's performance estimates were significantly lower than those for Crews 1 and 3 (mean performance index of .83 for Crew 2 *versus* .91 and .95 for Crews 1 and 3, respectively). For the task "Traverse Tube," performed by the Gunner, the performance estimates given by Crew 2 were again significantly lower than those given by Crews 1 and 3 (mean performance index of .61 for Crew 2, *versus* .74 and .76 for Crews 1 and 3, respectively).

For the task "Check Sight Picture, Level Bubbles," performed by the Gunner, Crew 3 gave significantly higher performance estimates than did Crews 1 and 2 (mean performance index of .74 for Crew 3 *versus* .63 for both Crews 1 and 2). For the task "Set Elevation on Range Quadrant," performed by the Assistant Gunner, Crew 2 gave performance estimates significantly lower than those of Crews 1 and 3 (mean performance index of .63 for Crew 2 *versus* .73 and .76 for Crews 1 and 3, respectively).

For the task "Elevate Tube," performed by the Assistant Gunner, Crew 2 again gave performance estimates significantly lower than those of Crews 1 and 3 (mean performance index of .56 for Crew 2 *versus* .69 and .70 for Crews 1 and 3, respectively). For the task "Load Projectile and Propellant," performed by the Cannoneers, Crew 3 gave significantly higher

performance estimates than did Crew 1 (mean performance index of .73 for Crew 3 *versus* .63 for Crew 1; Crew 2's mean [.69] did not differ from those of either Crew 1 or Crew 3).

Having found statistically-reliable differences among the performance estimates provided by members of the three crews for some tasks, we wished to assess these differences for *practical* significance. To do so, we computed the covariance statistic η^2 (Cohen, 1969) for each of the six analyses where differences were found between the three crews' estimates of performance. η^2 represents the proportion of total sample variance accounted for by membership in particular *a priori* groupings (in this case, the three crews), and is analogous to a correlation coefficient. The obtained values for η^2 are shown in Table 2-2. The obtained values for η^2 are all rather low, indicating that the differences found between the data from the three crews are probably of little concern.

Table 2-2. Values of the η^2 covariance statistic for six tasks with differences among means.

Task Title Corresponding to Analysis	η^2 Statistic
Receive and Relay Fire Order	.383
Traverse Tube	.190
Check Sight Picture, Level Bubbles	.368
Set Elevation on Range Quadrant	.266
Elevate Tube	.317
Load Projectile and Propellant	.216

Based on these findings, we decided to pool the data from the three crews for the purpose of computing regression equations describing the relationships between SME performance estimates on the tasks and time in MOPP 4. Therefore, all available data were used to compute the regression equations, without regard to the issue of pre- *versus* post-exercise administration of the questionnaires or to crew membership.

The regression equations were computed, task by task, in the normal fashion, using the logarithm of the performance indices as the predicted variable and time in MOPP 4 (in hours) as the sole predictor. The logarithm of the performance indices was used to permit direct comparison between the terms of regression equations based on SME-estimate data and those derived from the measured performance data. (The regression equations for measured performance data were prepared in a separate analysis [McClellan, Deverill, and Matheson, 1994] that also used the logarithms of performance indices.)

Summary statistics for pooled (across crew and pre- *versus* post-administration) SME-estimate performance index data at 1, 2, and 4 hours in MOPP 4 are presented in Appendix B.

SECTION 3
MEASURED PERFORMANCE DATA

As summarized in an earlier section of this report, a team of observers gathered task performance data from three howitzer crews performing field artillery fire missions both in BDU (Baseline condition), and in MOPP 4. Three observers normally performed the observations, each recording the times of specific, predefined events during each fire mission by pressing coded keys on the keyboards of notebook computers. Occasionally a fourth or fifth observer was used to check on the accuracy and consistency of data gathered by the three primary observers. For purposes of synchronizing the event records, all observers recorded the time of firing of each round during every fire mission. Details of the data collection method are found in McClellan (1992).

Ultimately, the event-timing data were reduced to performance times for several howitzer crewmember tasks, including either the exact tasks or close analogues of the tasks (see the discussion of task comparability later in this section) for which SME estimate data were obtained. A comprehensive discussion of the procedures for event time reconciliation, correcting for differences between event timing resulting from the drift of internal computer clocks, and other details of data reduction, is found in McClellan, Matheson, and Deverill (1994).

For comparison with other bodies of data similarly scored, as well as for the purposes of this validation effort, the raw task-time data were converted to performance indices. For each task, this conversion was performed by a self-normalization to Baseline task times, rather than using the (more or less) arbitrary nominal task times that were provided for reference in the SME performance estimation portion of the effort. This was accomplished by examining distributions of task times in the Baseline condition, removing gross outliers, and computing the mean task time for each task. The Baseline task time *means* were then used as reference times in the computation of performance indices for the same tasks in MOPP 4. The performance indices were computed by dividing the Baseline reference time (t_0) for a task by the observed MOPP 4 task completion time (t_i). Due to the statistical distributions of the measured data (Baseline and MOPP 4 conditions), this computation occasionally resulted in performance indices greater than 1.0 (*i.e.*, measured task performance time in MOPP 4 was less than the mean task performance time in the Baseline condition). Such data points are statistically valid and are retained, although they represent a conceptually confusing situation of "better than perfect" performance in the context of the performance index concept used in the SME performance estimation method, where statistical variations are not explicitly considered.

The measured performance-index data for each task were then used to derive regression equations relating task performance to time in MOPP 4 (the logarithm of the performance index was actually used in the computations). Since the times of fire missions relative to donning MOPP 4 equipment varied from observational day to day, and since crews were unable to accomplish some planned fire missions at all (due to crew size reductions during observational days), data corresponding exactly to the SME-estimate times in MOPP 4 do not exist. To provide a basis for comparisons of performance at these specific times, the regression equations

based on measured performance were used to obtain estimates of the mean, standard error of the mean, and standard deviation of performance for tasks at 1, 2, and 4 hours in MOPP 4. These data form the first portion of the comparison between SME-estimate and measured data in the following section. A comparison of the slope components of the regression equations for the tasks follows the detailed, time-by-time and task-by-task comparison.

3.1 TASK COMPARABILITY.

For the most part, the observational data collection methods discussed above were designed to capture data on the same tasks for which SMEs provided questionnaire-based performance estimates. Specific, observable events consisting of visual and/or auditory cues that reliably occur during the conduct of howitzer fire missions were selected as the bases for observational data collection. This maximizes the potential reliability and validity of the observational data.

Unfortunately, we cannot be sure that the performance estimates given by SMEs embody the same levels of reliability and validity as the data obtained by observation. In order to make a reasonably accurate performance estimate using the questionnaire method, an SME must possess a *mental model* of a task that encompasses a starting point, an ending point, and a concept of task duration. Further, the SME must possess or must create (in the estimation process) several different instantiations of this mental model, each corresponding to the conditions for which performance estimates are desired. "True" SMEs (see Section 1) are the most likely to have such a 'sheaf' of appropriate mental models to use in performance estimation and, therefore, to provide reasonably reliable and valid performance estimates. As noted, the personnel who performed in the SME role in this effort do not meet the characteristics of "true" SMEs. This means that it is not certain whether these personnel possessed valid baseline mental models of the tasks for which they (nevertheless) gave performance estimates for various conditions of time in MOPP 4.

The characteristics and administration of the questionnaire method may also have contributed to inaccurate mental model development on the part of SMEs that could weaken the comparability of SME and "measured" performance estimates. Some of the tasks for which performance estimates were given are performed differently during different variants on howitzer fire missions.

For example, use of different elevation angles of the cannon tube results in two variants of fire missions: "normal" and high angle. In a high angle fire mission, the tube's elevation must be lowered in order to load a subsequent round for firing (the loading tray for the projectile cannot be attached to the howitzer when elevation angles greater than 1 radian are used). This means that at least two additional steps are added to the task "Elevate Tube" when a high angle mission is fired: lowering the tube for the reloading process; and re-elevating the tube to the appropriate angle for firing (and assuring that the correct angle has in fact been achieved).

The instructions to SMEs for the estimation task were tacit on the issue of the mission variant hypothetically being fired, as well as on some other variable conditions that could

influence specifics of mental models developed by the SMEs in the estimation process. Since the measurement of mental models is problematic at best (see Rouse, 1991), there is no way to ascertain whether the instructions given produced the desired mental models of tasks. Inferentially, SMEs gave generally consistent estimates of performance that conformed to our expectations (*e.g.*, longer task performance times were generally estimated for longer periods in the MOPP 4 ensemble; SMEs did not typically provide "impossible" values for task performance time estimates, etc.). Therefore, some confidence can be placed in the quality of SMEs' task performance estimates, but caution is warranted in interpreting the results that follow.

Another issue with respect to SME estimates, also related to the ability to form appropriate mental models, is the amount and (for "pre-" estimates, unknown) recency of SMEs' performing firing section tasks while actually wearing the MOPP 4 ensemble. The available data (see Table 2-1 and Appendix A) suggest a relatively low level of SME experience even wearing MOPP 4, as well as (unquantifiable) limited experience in performing firing section tasks while in MOPP 4. Again, this is simply an issue to keep in mind when considering the comparison of SME-estimate with measured data in the following section of this report.

3.2 TASKS COMPARED.

While the observational data collection procedures were designed to capture data on howitzer crewmember tasks analogous to those used in the SME performance estimation portion of the work, some modest differences were found between the two task sets after the fact. These differences reflect the reality of firing actual fire missions, contrasted with the artificial and abstracted character of the tasks considered by SMEs in the estimation process. Two specific cases are relevant to the remainder of this report.

The first case involves the task "Load Projectile and Propellant," performed by the No. 1 Cannoneer with assistance from other Cannoneers. During data reduction, it was noted that the observable events and timing intervals used for timing this task were consistently longer for the first round of a fire mission than for subsequent rounds (*i.e.*, reloading is faster, based on the observable phenomena used in data collection). The decision was made to treat the tasks associated with the first rounds of fire missions as a different task than those associated with subsequent rounds of fire missions. Both these tasks (first round and reload) are compared to the undifferentiated SME estimation task "Load Projectile and Propellant," in the analyses reported in the next section.

The second case involves the task "Open Breech, Inspect Bore," also performed by the No. 1 Cannoneer. Here, the task associated with the *last* round of a fire mission differs from the task of the same name performed in conjunction with previous rounds. For rounds other than the last, the No. 1 Cannoneer uses a wet swab to remove propellant residue from the howitzer's firing chamber prior to loading the next round. This was noted by the observers to be only a swift swabbing of the firing chamber, for all rounds save the last. After the final round of a fire mission, a more thorough swabbing of the firing chamber and breech block takes place. Since this more

thorough swabbing after the last round requires more time than that for previous rounds, the last-round data were excluded from the analysis to derive a regression equation for this task.

Two timed components of this task ("Open Breech, Inspect Bore") were measured in the observational data collection process: physically unlatching and opening the breech; and swabbing the firing chamber. The derived performance indices for these two components are compared separately with the SME-estimates for this task in the following analyses. The reason for this is that we are unsure which component or components the SMEs included in their mental models of task performance when making performance estimates.

For the analyses presented in the next section, the SME-estimate and corresponding observationally-measured tasks shown in Table 3-1 were used.

Table 3-1. SME-estimate and observationally-derived task comparisons.

Crewmember Performing Task	Observationally-Measured Task	SME-Estimation Task
Chief of Section	Receive and Relay Fire Order	Receive and Relay Fire Order
Gunner	Set Deflection on Sight	Set Deflection on Sight
Gunner	Traverse Tube (includes fine adjustment)	Traverse Tube and Level Bubbles
Gunner	Check Sight Picture and Level Bubbles	Check Sight Picture and Level Bubbles
Assistant Gunner	Set Elevation on Range Quadrant	Set Elevation on Range Quadrant
Assistant Gunner	Elevate Tube (includes fine adjustment)	Elevate Tube and Level Bubbles
Cannoneers	Load Projectile and Propellant (First Round)	Load Projectile and Propellant
Cannoneers	Load Projectile and Propellant (Reload)	Load Projectile and Propellant
No. 1 Cannoneer	Lock Breech and Prime Firing Mechanism	Close Breech and Prime Firing Mechanism
No. 1 Cannoneer	Open Breech (Open)	Open Breech and Inspect Bore
No. 1 Cannoneer	Swab Chamber (Swab)	Open Breech and Inspect Bore

3.3 TASK BASELINE TIMES.

As mentioned earlier, SMEs were given doctrinally-approved baseline or nominal times for howitzer crewmember task performance to consider in making task performance estimates. In contrast, measured task times were used as Baseline times for tasks in computing the performance indices for the measured data set used here. While the normalization inherent in computing performance indices should make the issue of different baseline times a moot one, there were some observed differences between the SME "baseline" times and baseline times derived from observation. This subsection presents task baseline times and points out the differences between those used in the SME estimation process and those derived empirically.

Table 3-2 shows the baseline times used for developing performance indices for the SME-estimated and the corresponding observationally-derived tasks. For the observationally-derived tasks, the mean and the interval of one standard deviation above and below the mean are given. (The plus and minus standard deviation values are asymmetric about the mean, since they are derived from regression equations developed using the logarithms of observed task times.) It is interesting to note that, in many cases, the empirically-derived baseline task times are longer than the nominal times used in the SME estimation process. The most extreme cases of this involve the tasks "Traverse Tube," performed by the Gunner, and "Elevate Tube," performed by the Assistant Gunner. These tasks, as observed, included fine adjustments to bring the azimuth and elevation of the cannon exactly in line with the specifications in the fire order. This was presumably also the case for the SME-derived performance estimates, since the description of each of the tasks included the phrase "level bubbles," which is shorthand for making the fine adjustments required.

This brings into question the derivation of the doctrinal task times, including the issue of the appropriateness of the mental models of the tasks possessed by those who established these nominal task times in the first place. Alternatively, it may be the case that the howitzer crews who participated in this research are less expert (and, therefore, slower) than (hypothetical) personnel (hypothetically) observed to establish the doctrinal times. The issue is not one that can be resolved here, but it should be kept in mind for future uses of the SME-cum-questionnaire task performance assessment method.

Table 3-2. Baseline times (seconds) used in developing performance indices for tasks.

Crewmember Performing Task	Observationally-Measured Task	Measured Baseline Time (Mean; ± 1 s.d.)	SME-Estimate Baseline Time
Chief of Section	Receive and Relay Fire Order	25.0 (+5.0; -4.0)	25
Gunner	Set Deflection on Sight	5.4 (+2.2; -1.5)	4
Gunner	Traverse Tube	18.1 (+5.4; -4.0)	7
Gunner	Check Sight Picture and Level Bubbles	5.5 (+2.9; -1.9)	4
Assistant Gunner	Set Elevation on Range Quadrant	6.8 (+2.4; -1.8)	4
Assistant Gunner	Elevate Tube	19.6 (+6.0; -4.6)	8
Cannoneers	Load Projectile and Propellant (First Round)	9.8 (+3.2; -2.4)	8
Cannoneers	Load Projectile and Propellant (Reload)	7.5 (+1.3; -1.1)	8
No. 1 Cannoneer	Close Breech and Prime Firing Mechanism	4.0 (+1.5; -1.1)	5
No. 1 Cannoneer	Open Breech and Inspect (Open)	2.1 (+0.5; -0.4)	5
No. 1 Cannoneer	Open Breech and Inspect (Swab)	5.0 (+0.9; -0.7)	5

SECTION 4
COMPARISON METHODS AND RESULTS

Two types of comparisons between the SME-estimate task performance data and the observationally-derived task performance data were made.

The first comparison contrasts the performance estimates from the two sources task-by-task at three different times in MOPP 4: 1 hour, 2 hours, and 4 hours. The data points used for these comparisons are the mean and standard deviation statistics for the SME-derived and observationally-derived task performances, in terms of the logarithms of performance metrics. The comparison was performed using only summary statistics because actual measured data points for the observationally-derived data do not correspond exactly to the times in MOPP 4 for which SMEs gave performance estimates. Rather, the statistical parameters for the observational data were calculated from the regression equations describing the relationship between performance and time in MOPP 4 for these data. The summary statistics for the SME-derived data were computed directly from SME estimates at each of the three times in MOPP 4. Task-by-task comparisons between the SME estimate data and the observationally-derived data were made by Student's t-test (Dixon and Massey, 1957). Due to the often extremely discrepant ns for the samples, as well as frequently severe heterogeneity of variance between samples, pooled error terms were computed for all t-statistics reported below.

The astute reader will note the use of fractional sample sizes and degrees of freedom used for some observationally-derived data in the tables in the next subsection. This is due to the use of performance estimate statistics derived from the regression equations for observationally-derived task performance. Normally, raw data are desirable for computing the comparison t-statistics, since a definite number of degrees of freedom can be used in selecting the critical value for assessing the statistical significance of the test. In this case, however, no definite number of degrees of freedom was associated with each of the observationally-derived data points. In order to perform the comparison t-tests, we simply divided the number of observations contributing to the derivation of each of the regression equations by 3, and used the resulting number (whether integer or fraction) as the n for computing the tests.

The second comparison between the observationally-derived and SME-estimated data contrasts the slope terms of the regression equations derived to predict performance as a function of time in MOPP 4. Comparing the intercept terms of the equations is not meaningful in this case, since these terms represent a questionably valid projection of measured performance at a "time zero" with respect to wearing the MOPP 4 ensemble. No data were taken immediately after donning MOPP 4. The SME-estimate data points for 5 minutes in MOPP 4 are excluded from computation of the regression equations for SME-estimated data in order to provide a more directly comparable set of equations. As noted above, the equations are computed using the logarithms of the performance indices. Therefore, each equation is of the form

$$\log(\text{performance index}) = \text{constant} + (\text{slope} \times \text{Time in MOPP 4 [hours]}).$$

The equation terms (and standard errors used in computation) are included in Table 4-5 in Section 4.2 below.

Standard errors were available (from computation of the regression equations) for both constant and slope terms. Since the slope terms were universally fractions, we chose to use a statistical test of the difference in two proportions (Guilford, 1956) for comparing them. This test yields a statistic interpretable as a z-score.

4.1 RESULTS: TASK-BY-TASK AND TIME-BY-TIME PERFORMANCE COMPARISONS.

Tables 4-1, 4-2, and 4-3 on the following pages summarize the results of the comparisons between SME-estimate and observationally-derived performance indices for each of the 11 task comparisons at 1, 2, and 4 hours in MOPP 4, respectively. These tables present the summary statistical data (mean, sample size, standard error of the mean, and standard deviation) for the two samples in parallel rows for each task. The summaries for the observationally-derived samples are in the rows labeled M in the second column of the tables; SME-estimate derived data summaries are in rows labeled S. This information is followed in the rightmost three columns by results of the t-tests performed to compare the samples: computed t statistic, degrees of freedom, and level of statistical significance achieved by the test.

It is clear from the results in Tables 4-1, 4-2, and 4-3 that the SMEs responding to the performance-estimation questionnaires tended to overestimate performance deterioration due to MOPP 4 somewhat more frequently than they underestimated. That is, the performance indices from SME-estimates of performance were less than those derived from measured data in more cases (19 of 29, or about 65 percent) than they were greater (6, or about 21 percent). In four cases, the estimates from SME and measured data agreed closely.

Fifteen of the 29 t-tests (about 52 percent) achieved statistical significance at the 95 percent level of confidence or greater. Of these, 11 were cases where SME-estimates of performance were less than estimates derived from observational data (73 percent of cases where significance was reached); the remaining 4 (27 percent) were cases where SME performance estimates were larger than observationally-derived performance indices. Of the four cases where SME estimates of performance were significantly higher than those from observational data, three were associated with one task: "Receive and Relay Fire Order," performed by the Chief of Section.

Table 4-1. Comparison of field-measured data and SME estimates after 1 hour in MOPP 4.

Task[1]		Log (Performance Index) Statistics				Comparison Statistics		
		Mean	n	s_x	s	t	df	p
Rec. & Relay Fire Order	M[2]	-.119	4	.047	.138	2.18	61	<.025
	S[3]	-.038	59	.009	.067			
Set Deflection on Sight	M	-.186	4.33	.078	.230	0.37	60.33	>.10
	S	-.153	58	.023	.173			
Traverse Tube	M	-.225	4.5	.029	.068	0.93	59.5	>.10
	S	-.156	57	.021	.156			
Check Sight	M	-.378	4.33	.099	.304	2.04	60.33	<.025
	S	-.184	58	.024	.182			
Set Elevation on Quadrant	M	-.121	5.5	.039	.089	0.35	61.5	>.10
	S	-.148	58	.023	.176			
Elevate Tube	M	-.030	5	.043	.106	2.29	61	<.025
	S	-.190	58	.020	.152			
Load Projo/ Powder (First)	M	-.033	4.33	.037	.110	1.58	62.33	>.05
	S	-.168	60	.023	.175			
Load Projo/ Powder (Rel'd)	M	-.085	12.66	.011	.059	1.67	70.66	<.05
	S	-.168	60	.023	.175			
Close Breech & Prime	M	-.139	25	.026	.114	0.09	80	>.10
	S	-.136	57	.020	.154			
Open Breech & Inspect (Open)	M	-.029	15.33	.017	.089	1.97	72.33	<.05
	S	-.100	59	.017	.134			
Open Breech & Inspect (Swab)	M	-.097	15.33	.013	.071	0.08	72.33	>.10
	S	-.100	59	.017	.134			

[1] M indicates measured data; S indicates SME estimate data
[2] Values predicted at 1 H in MOPP 4 from regression analysis of measured data
[3] Values calculated directly from SME estimates for 1 H in MOPP 4

Table 4-2. Comparison of field-measured data and SME estimates after 2 hours in MOPP 4.

Task[1]		Log (Performance Index) Statistics				Comparison Statistics		
		Mean	n	$s_{\bar{x}}$	s	t	df	p
Rec. & Relay Fire Order	M[2]	-.169	4	.036	.135	2.58	59	<.01
	S[3]	-.065	57	.010	.074			
Set Deflection on Sight	M	-.112	4.33	.058	.223	1.03	60.33	>.10
	S	-.212	58	.025	.193			
Traverse Tube	M	-.098	4.5	.023	.065	1.43	61.5	>.05
	S	-.212	59	.022	.167			
Check Sight	M	-.242	4.33	.077	.296	0.06	60.33	>.10
	S	-.248	58	.026	.197			
Set Elevation on Quadrant	M	-.035	5.5	.027	.084	2.15	60.5	<.025
	S	-.222	57	.027	.201			
Elevate Tube	M	-.028	5	.036	.103	2.95	58	<.005
	S	-.264	55	.024	.175			
Load Projo/ Powder (First)	M	-.059	4.33	.028	.107	1.63	60.33	>.05
	S	-.271	58	.035	.267			
Load Projo/ Powder (Rel'd)	M	-.133	12.66	.010	.058	1.82	68.66	<.05
	S	-.271	58	.035	.267			
Close Breech & Prime	M	-.246	25	.026	.114	0.37	79	>.10
	S	-.232	56	.023	.175			
Open Breech & Inspect (Open)	M	-.062	15.33	.013	.088	2.29	73.33	<.025
	S	-.169	60	.023	.176			
Open Breech & Inspect (Swab)	M	-.142	15.33	.010	.070	0.59	73.33	>.10
	S	-.169	60	.023	.176			

1 M indicates measured data; S indicates SME estimate data
2 Values predicted at 2 H in MOPP 4 from regression analysis of measured data
3 Values calculated directly from SME estimates for 2 H in MOPP 4

Table 4-3. Comparison of field-measured data and SME estimates after 4 hours in MOPP 4.

Task[1]		Log (Performance Index) Statistics				Comparison Statistics		
		Mean	*n*	s_x	*s*	*t*	df	*p*
Rec. & Relay Fire Order	M[2]	-.267	4	.081	.155	3.30	58	<.005
	S[3]	-.098	56	.013	.095			
Set Deflection on Sight	M	+.037	4.33	.132	.255	2.85	57.33	<.005
	S	-.320	55	.034	.251			
Traverse Tube	M	No Data Available				No Comparison Possible		
	S	-.320	58	.028	.214			
Check Sight	M	+.030	4.33	.179	.341	3.14	58.33	<.005
	S	-.354	56	.032	.238			
Set Elevation on Quadrant	M	No Data Available				No Comparison Possible		
	S	-.333	56	.034	.256			
Elevate Tube	M	No Data Available				No Comparison Possible		
	S	-.370	58	.029	.217			
Load Projo/ Powder (First)	M	-.114	4.33	.064	.123	1.97	61.33	<.05
	S	-.326	59	.030	.220			
Load Projo/ Powder (Rel'd)	M	-.228	12.66	.022	.062	1.57	69.66	>.05
	S	-.326	59	.030	.220			
Close Breech & Prime	M	No Data Available				No Comparison Possible		
	S	-.318	57	.031	.234			
Open Breech & Inspect (Open)	M	-.129	15.33	.032	.093	1.77	70.33	<.05
	S	-.256	57	.031	.230			
Open Breech & Inspect (Swab)	M	-.234	15.33	026	.072	0.37	70.33	>.10
	S	-.256	57	.031	.230			

1 M indicates measured data; S indicates SME estimate data
2 Values predicted at 4 H in MOPP 4 from regression analysis of measured data
3 Values calculated directly from SME estimates for 4 H in MOPP 4

At 1 hour in MOPP 4, SMEs overestimated performance deterioration (5 cases, or about 46 percent) relative to observational data about as frequently as they underestimated (4 cases, or about 36 percent). At 1 hour in MOPP 4, 5 of the 11 by-task comparisons reached statistical significance. In three of the cases where *t*-tests achieved significance (60 percent), SMEs estimated more performance deterioration than was found in the measured data; in the other two cases, SMEs underestimated performance change.

At 2 hours in MOPP 4, SME overestimation of performance deterioration relative to observational data was more pronounced. SMEs estimated lower levels of performance than reflected in the observational data for 8 tasks (about 73 percent), and higher levels for only 1 task (about 9 percent). In the remaining two cases, SME performance estimates were almost the same as the data from observation. At 2 hours in MOPP 4, five of the 11 task-by-task comparisons reached statistical significance. In four of these cases (80 percent), SME performance estimates were lower than observational data; for the remaining case, the SME estimate was higher.

At 4 hours in MOPP 4, SMEs overestimated performance deterioration for 6 of the 7 tasks, (about 86 percent) for which comparisons were possible (there were no observational data available at 4 hours in MOPP 4 for four of the tasks). In the remaining case, SMEs underestimated performance deterioration. Five of the 7 *t*-tests for task-by-task comparison at 4 hours in MOPP 4 were statistically significant. In four of these cases, or 80 percent, SMEs overestimated performance deterioration; in the fifth case, SMEs estimated less performance deterioration than was reflected by observational data.

From these data, it is reasonable to state that SMEs have a tendency to overestimate performance deterioration as a function of time in MOPP 4, relative to measured performance data. In addition, the proportion of cases where SMEs overes[illegible]e performance change appears to increase with the amount of time in MOPP 4 for which estimates are given. For 1 hour in MOPP 4, SMEs overestimated performance change for about 45 percent of tasks; for 2 hours in MOPP 4, the percentage of tasks where SMEs overestimated change increased to 73 percent; and for 4 hours in MOPP 4, the percentage rose to about 86 percent. For cases where the comparisons of SME estimates and observationally-derived performance measures reached statistical significance, there is a similar trend. The proportion of cases where significant *t*-tests were found, and SMEs overestimated performance change, increased from 60 percent at 1 hour in MOPP 4 to 80 percent at 2 hours and 4 hours in MOPP 4.

Turning our attention now to the results with respect to specific tasks, we find the following:

- For only one task did SMEs consistently *underestimate* the amount of performance change relative to observationally-derived performance data. This was for the task "Receive and Relay Fire Order," performed by the Chief of Section. At all three times in MOPP 4 for which SME estimates were made, SMEs underestimated the performance change found in observational data for this task (all three comparisons were statistically significant). Further, the absolute difference between SME estimates and observational data increased as a function of time in MOPP 4 for which

estimates were made: 17 performance-index points of SME underestimate at 1 hour, 19 points at 2 hours, and 28 points at 4 hours.

- For several other tasks, SMEs consistently *overestimated* performance change. For the task "Load Projectile and Propellant (First Round)," performed by the Cannoneers, SMEs' estimates of task performance deterioration were greater than those shown in the observational data at all three estimate times; the comparison at 4 hours in MOPP 4 was statistically significant. For the task "Load Projectile and Propellant (Reload)," performed by the Cannoneers, SMEs' estimates of task performance deterioration were consistently greater than observational data, and statistical significance was achieved for the comparison tests at 1 and 2 hours in MOPP 4. For the task "Open Breech and Inspect (Open)," performed by the No. 1 Cannoneer, SMEs overestimated performance change at all three estimate times, the comparisons reaching statistical significance at all three times in MOPP 4. For the task "Elevate Tube," performed by the Assistant Gunner, SMEs overestimated performance change at both 1 and 2 hours in MOPP 4; comparisons at both times were statistically significant.

- For the remaining tasks, SMEs sometimes overestimated and sometimes underestimated performance change relative to observational data. The general tendency with these tasks appears to have been for SMEs to somewhat underestimate performance change at the 1-hour estimate and to somewhat overestimate performance change at longer estimate intervals. This is by no means consistent across tasks, however, but merely an apparent trend with a majority of the remaining tasks.

To aid in understanding these mixed results, a supplemental analysis was done. In this analysis, we compare the rated demand of the nine SME-estimation tasks on five major human abilities with an artificial variable that reflects SMEs' tendency to over- or under-estimate performance change. The artificial variable is defined to evaluate the net trend toward over- or under-estimation of performance change for each of the nine tasks. We simply counted the number of cases per task where SMEs had underestimated performance change at the three estimate times, and subtracted from the total the number of cases where SMEs had overestimated. This "estimation" variable can range in integer value from +3 (consistent underestimation of performance change) to -3 (consistent overestimation of performance change).

The estimation variable was correlated with the tasks' rated demands on five abilities: Attention, Perception, Psychomotor, Physical, and Cognitive (see Roth, 1991, 1992, and Anno, Dore and Roth [in preparation] for more detail on the ability demand rating process), as well as with an overall ability demand rating (just the sum of tasks' individual ratings on the five discrete abilities). Ability demand ratings used are those developed for an analogous set of howitzer crewmember tasks in the AQUAFHL effort (Anno, Dore, Roth, LaVine, and Deverill, 1994). The data used for this analysis and the resulting correlation coefficients are presented in Table 4-4. Ability demand ratings can range from 1 (low demand) to 7 (high demand for an ability).

Table 4-4. Comparison of estimation variable with rated ability demands for 9 tasks.

Task	Value of Estimation Variable	Rated Demand on Five Human Abilities					Overall Ability Demand Rating
		Attention	Perception	Psychomotor	Physical	Cognitive	
Receive and Relay Fire Order	+3	4	2.5	1	1	4	12.5
Set Deflection on Sight	-1	5	3	4	2.5	4.3	18.8
Traverse Tube	0	5	3	4	2.5	4.3	18.8
Check Sight Picture, Level Bubbles	0	5	3	4	2.5	4.3	18.8
Set Elevation on Range Quadrant	-2	5	3	4	2.5	4.3	18.8
Elevate Tube	-2	5	3	4	2.5	4.3	18.8
Load Projectile and Propellant	-3	4.3	2.5	3.3	6	2.3	18.4
Close Breech and Prime	0	4.3	3	3.3	1.8	3.3	15.7
Open Breech and Inspect	-2	3.3	4	2.8	3.3	2.8	16.2
Correlation Coefficient		+.079	-.323	-.621	-.760	+.395	-.700

Only one of the correlation coefficients between the estimation variable and rated demand for a specific ability achieved statistical significance at the 95 percent level of confidence. This is the correlation between tasks' rated *physical* demand and the estimation variable, at -.760. The interpretation of this correlation coefficient is that SMEs consistently *underestimated* performance change due to wearing the MOPP 4 ensemble for tasks rated *lower in physical demand*, and consistently *overestimated* performance change for tasks rated as *higher in physical demand*. SMEs thus overestimate the effects on task performance of enclosure in the MOPP 4 ensemble for more physically-demanding tasks.

None of the other correlation coefficients for specific abilities reached statistical significance (the [nondirectional] critical value of r for 8 degrees of freedom and α=.05 is .632). The achieved correlations do suggest some tendencies, however. It appears that SMEs may have a tendency to underestimate performance change for tasks rated lower in demand for psychomotor abilities. Conversely, SMEs may overestimate performance change for tasks rated higher in demand for psychomotor abilities.

The correlation (-.700) between the estimation variable and the overall ability demand rating achieved statistical significance. We interpret this to mean that SMEs have a tendency to *over*estimate performance decrements associated with being in MOPP 4 for tasks that are

perceived to be more demanding or effortful, overall, and to *under*estimate performance decrements for less-effortful tasks.

In summary, SMEs' estimates of performance change due to enclosure in the MOPP 4 ensemble are generally similar to performance change data derived from observation. However, in point-by-point comparison, some interesting differences arise. In about half the cases where comparisons were made, SME performance change estimates differed statistically from observationally-derived data. SMEs overestimated performance change more frequently than they underestimated. In one case (the task "Receive and Relay Fire Order"), SMEs *under*estimated performance change for all three periods in MOPP 4, the underestimate increasing in magnitude with longer time in MOPP 4. For several other tasks, SMEs tended to consistently *over*estimate performance change due to MOPP 4 conditions, across the three estimation times. There was some tendency for the amount of the overestimation of performance change to increase for longer times in MOPP 4 for these tasks, but this was by no means consistent. SMEs' tendency for under- *versus* over-estimation of performance change was related to the rated demand for physical ability of tasks, and to the overall ability demand of tasks. SMEs overestimate performance change due to MOPP 4 enclosure for tasks that inherently demand larger amounts of physical (or overall) ability, and underestimate performance change for less physically-demanding tasks and tasks that are less demanding, overall.

4.2 RESULTS: REGRESSION EQUATION SLOPE TERM COMPARISONS.

The results of the statistical comparison of the slope terms of the regression equations derived from observationally-based performance data and SME estimates of performance are shown in Table 4-5. Two rows appear in the table for each task. The top row, labeled "M" in the second column of the table, corresponds to the regression equation derived from observationally-based data. The lower row, labeled "S," corresponds to the equation for SME-estimate data. The third and fourth column entries in a row are the value of the regression equation intercept term and its standard error. (These are included only for completeness.) The fifth and sixth columns contain the value of the regression equation slope term and its standard error. In the rightmost column the results of comparing the slope terms of the observationally- and SME-derived equations are shown: the computed denominator term σ_{d_b} is shown in the top row; the derived *z*-statistic comparing the slope term values is in the lower, shaded row.

Supplementing the data presented in Table 4-5 are comparison plots of performance indices derived from SME-estimate data and observational data, shown in Appendix C. In these plots, observationally-derived data are represented by shaded areas. More darkly-shaded areas indicate the interval about the mean of ± one standard error of the mean (means are plotted symbolically using oval shapes). Lighter-shaded areas represent the interval about the mean of ± one standard deviation. SME-estimate data are represented by plotted symbols and graph lines connecting them. The mean is indicated in these plots by rectangular shapes with crossed lines inside the rectangles. Intervals of ± one standard error of the mean are denoted by outlined triangles. Intervals of ± one standard deviation are indicated by filled triangular shapes. Because

of the occasionally large error bounds of the estimates, some points are not within the ordinate scale of the plots, and are omitted.

Table 4-5. Comparison of SME-estimate and measured regression equation components.

Task	Src	Regression Constant		Regression Slope		
		Value	Standard Error	Value	Standard Error	σ_{dp}† / *z-statistic*
Rec. & Relay Fire Order	M	-.07009	.07401	-.04930	.03435	.03469
	S	-.02110	.12678	-.01968	.00483	0.850
Set Deflection on Sight	M	-.26032	.12401	+.07437	.05729	.05870
	S	-.09839	.03342	-.05573	.01278	2.216*
Traverse Tube	M	-.35222	.06163	+.12709	.03597	.03761
	S	-.10169	.02114	-.05461	.01100	4.831**
Check Sight	M	-.51353	.15934	+.13597	.07561	.07686
	S	-.13110	.03325	-.05620	.01266	2.507**
Set Elevation on Quadrant	M	-.20742	.08011	+.08646	.04544	.04728
	S	-.09285	.03430	-.06061	.01305	3.111**
Elevate Tube	M	-.03342	.09064	+.00298	.05387	.05506
	S	-.13637	.02961	-.05421	.01119	1.130
Load Projo/ Powder (First)	M	-.00552	.05823	-.02705	.02738	.02975
	S	-.13867	.03584	-.04986	.01382	0.718
Load Projo/ Powder (Rel'd)	M	-.03757	.01784	-.04770	.00874	.01635
	S	-.13867	.03584	-.04986	.01382	0.132
Close Breech & Prime	M	-.03303	.06327	-.10648	.04096	.04260
	S	-.09227	.03103	-.05810	.01171	1.135
Open Breech & Inspect (Open)	M	+.00463	.02710	-.03332	.01332	.01738
	S	-.05687	.02929	-.05061	.01117	0.995
Open Breech & Inspect (Swab)	M	-.05103	.02171	-.04567	.01066	.01544
	S	-.05687	.02929	-.05061	.01117	0.320

†Denominator term for testing the difference between two proportions
* $P < .05$
** $P < .01$

It is apparent from Table 4-5 that, for 7 of the 11 tasks, the rate of performance degradation with time in MOPP 4 (the slope terms of the regression equations) is statistically equivalent for the SME estimates and for field-measured data. Comparisons of the slope terms of the regression equations resulted in statistically significant differences for four of the 11 tasks. They are:

- Set Deflection on Sight, performed by the Gunner;
- Traverse Tube and Level Bubbles, performed by the Gunner;
- Check Sight and Level Bubbles, performed by the Gunner; and
- Set Elevation on Quadrant, performed by the Assistant Gunner.

For these tasks, SMEs estimated more or less monotonic decreases in performance with increasing time in MOPP 4. Observational data, however, indicate that performance actually *improved* with increasing time in MOPP 4. For one other task (Elevate Tube and Level Bubbles, performed by the Assistant Gunner), a similar trend was observed in the observational data, but the comparison of regression equation slopes did not achieve statistical significance. In the remaining six tasks' regression equations, both SME estimates and observationally-derived data indicate a monotonic decrease in performance with increasing time in MOPP 4, and the slope terms of the equations associated with the tasks do not differ by statistical comparison.

This result of *increasing* levels of task performance with increasing time in MOPP 4 is both counterintuitive and puzzling. Several considerations may help to explain this result.

First, the experience levels of crewmembers participating in the exercise were probably not high, as reflected by their time in service (see Table 2-1 and Appendix A). This could mean that the crewmembers occupying the Gunner and Assistant Gunner positions were still learning the component skills needed to perform the tasks where performance increases were found. An examination of the comparison plots in Appendix C shows that the means of the performance indices for the three tasks performed by the Gunner were considerably lower at 1 hour in MOPP 4 than SME-estimated performance at the same time in MOPP 4. By 2 hours in MOPP 4, however, observed performance on these tasks more closely approximated SMEs' performance estimates, as well as showing an absolute increase in performance relative to measured performance at 1 hour in MOPP 4 (*i.e.*, decreased time to perform the tasks). This supports the hypothesis that continued learning of the component skills involved in performing the tasks may have been occurring, even under conditions (MOPP 4) where performance deterioration would be expected.

This speculation is reinforced to some extent by an examination of the rated demands of these tasks for various human abilities (that underlie performance; see Table 4-4). The five tasks for which performance *increases* are observed are each rated higher than the remaining six tasks on demand for the abilities Attention, Psychomotor, and Cognitive. Skills supporting the required abilities in at least two of these domains (Psychomotor and Cognitive) tend to require many repetitions of a task or task element before mastery of the skills is attained (Bilodeau and Bilodeau, 1969). Also, tasks requiring significant amount of focused Attention for proficient

performance may require skill development over many practice sessions (Mackworth, 1970). Since the crewmembers performing the Gunner's and Assistant Gunner's tasks were most likely inexperienced, especially at performing the tasks in MOPP 4, acquisition of the underlying skills required for task performance was probably still taking place during the exercise. This is reflected by the absolute improvements in task performance from 1 hour in MOPP 4 to 2 hours in MOPP 4.

Note also from Table 4-4 that the five tasks are not rated as being highly physically demanding. Therefore, they may not have been affected to as great an extent by MOPP 4 conditions (*e.g.*, encumbrance, heat stress) as were more physically-demanding tasks. Thus, effects of increased proficiency through learning and repetition of tasks may have overshadowed the performance-degrading effects of enclosure in MOPP 4, for these tasks.

A possible argument against this interpretation is that, during the course of conducting fire missions, crewmembers were sometimes removed from active participation by the overseeing physician when their heart rates or core temperatures reached maximum safe limits. In some cases, this could result in substituting a crewmember with less experience in performing particular tasks for one with more experience (even very recently-gained experience) and, presumably, with greater proficiency. This could induce lower levels of performance of some tasks at longer times in MOPP 4, thus leveling-out or even reversing the apparent trend toward improved performance at longer times in MOPP 4. This does not appear to have been the case in this effort, although such a conclusion must be tentative since data were not specifically collected to address this issue.

An alternate interpretation, at least for some of these tasks, is that few or no data were gathered for time in MOPP 4 as long as 4 hours. In most cases, this was due to crewmembers being removed from participation in the exercise due to attaining maximum safe heart rate or core temperature limits. When crew size was reduced below seven, a crew's participation for the day was terminated. This frequently took place well before four hours of performance had been attained by a crew.

For example, for the task "Traverse Tube and Level Bubbles," performed by the Gunner, no data for times in MOPP 4 beyond about 2-½ hours were included in computing the regression equation. There are similar limitations in the data for the tasks "Elevate Tube and Level Bubbles" and "Set Elevation on Range Quadrant," performed by the Assistant Gunner. For the task "Check Sight, Level Bubbles," performed by the Gunner, only one observation beyond 2-½ hours was included in computing the regression equation. The absence of data representing performance at longer times in MOPP 4 may thus present a picture of performance change that is not strictly comparable to SME-estimate data that *always* included estimation of performance at longer times in MOPP 4.

It should be noted that the SME-estimation method was developed for use in just such cases. When actual performance data cannot be gathered, or is incomplete due to hazards to personnel or other constraints, the SME-estimation methodology should be considered a standard means of obtaining data on performance change due to the presence of battlefield stressors. The results presented here reinforce the validity and value of the SME-estimation method.

SECTION 5
DISCUSSION AND CONCLUSIONS

From these results, we conclude that the questionnaire-based SME-estimate method for assessing performance change in response to stressor exposure is conditionally validated. While SMEs had a tendency to overestimate the effect on performance of enclosure in the MOPP 4 ensemble for physically demanding tasks and to underestimate the effect for physically undemanding tasks, in general the SMEs made predictions of performance change that are more or less accurate when compared to observationally-measured howitzer-crew task performance. These results were found despite the less-than-ideal characteristics of the SMEs who gave performance estimates, and some limitations in the observationally-derived task performance data.

We claim only conditional validity for the SME-estimation method because of these limitations and since the sample of tasks on which the results are based is small. Additional data, gathered using improved preparation and administration procedures for the questionnaire assessment process (see below), and including tasks with more extreme variation in the ability demands exerted by tasks on the task performers, could lead to stronger demonstrations of validity for this method. However, the present results are encouraging and support continued use of the questionnaire-based SME-estimation method.

Some improvements to this method, as used in this effort, are indicated. These are offered to strengthen the method in the direction of enabling SMEs to form accurate and appropriate mental models of task performance and the effects of the conditions for which performance estimates are desired.

First, "true" SMEs, meeting the criteria outlined earlier in this report, should be used to perform the estimation process. This will maximize the likelihood of obtaining data that are as reliable and valid as human limitations in forming mental models of tasks permit.

Second, the nominal or baseline times for tasks used in the estimation method should be based on measured task performance under typical (non-stressed) conditions, whenever possible. There were some considerable discrepancies between the baseline times found in the observationally-based data in this effort, and the nominal task times used in the performance-estimation questionnaires. While every effort was made to obtain accurate doctrinally-based information[1] on nominal task performance times, it is not known how the times actually used were derived. When nominal times cannot be based on measured performance, the best available estimates should be reviewed by an independent panel of "true" SMEs (not those who will provide the performance estimates) and corrections made to the nominal task times as indicated by their review.

---

[1]Nominal task times were provided by the Gunnery Department of the U.S. Army Field Artillery School at Fort Sill, OK. The nominal times are reportedly based on field data and data contained in Army Training and Evaluation Program (ARTEP) publications.

Third, when possible, complete and thorough descriptions of the tasks for which performance estimates are desired should be provided, rather than only the task titles, as has been done in the past. These descriptions should include specifics of which variants of tasks are under consideration, when variants are known to exist, and any other relevant and specific details that will foster the development of complete, accurate, and consistent (between SMEs) mental models of the tasks. While these descriptions probably should not be included on each page of the estimation questionnaire, they should be available for reference by the SMEs who perform the task. A detailed review of the task descriptions should precede the beginning of the estimation task by SMEs.

Finally, attempts to verify that all SMEs have developed complete, accurate, and consistent mental models of each task for which estimation is to be made would be of value in future applications of this method. The assessment of mental models is difficult and to some extent subjective at this point, so specific methods are not offered. However, as the state of the art in this aspect of cognitive science improves (see Rouse, 1991, for comments), advantage of the improvements should be taken to further improve the SME-estimation method.

Methods for gathering actual, measured performance data might also be improved in some meaningful ways. Of particular note is the problem in this work of terminating crewmembers' participation in the exercise due to reaching maximum safe levels of physiological parameters. Reductions in the amount of time crews could participate in the exercise on any given day because of crewmember "dropout" rates severely limited the amount of data available on performance at longer times in MOPP 4. Also, some remaining confounding between the effects of "pure" environmental stresses of summertime heat and humidity, and the heat- and encumbrance-stresses induced by MOPP 4 conditions, may be present in the measured performance data from this exercise. Although the performance-decrement data were normalized using task performance times measured in BDU conditions, it is not certain that the baseline measures taken would be identical to baseline measures taken under less inherently stressful conditions.

These issues might have been avoided or minimized if the exercise had taken place at a cooler and less humid time of year, or in a geographic location where more moderate seasonal conditions could be expected.

SECTION 6
REFERENCES

Anno, G.H., Dore, M.E., Roth, J.T., LaVine, N.D., and Deverill, A.P. (1994) Predicted performance of infantry and artillery personnel following acute radiation or chemical agent exposure. DNA-TR-93-174. Washington, D.C.: Defense Nuclear Agency, July, 1994. (U)

Anno, G.H., Dore, M.E, and Roth J.T. (in preparation). Taxonomic model for performance degradation in combat tasks. Santa Monica, CA: Pacific-Sierra Research Corporation, in preparation. (Prepared under Contract DNA001-90-C-0139, "Combined Arms Combat Crew Performance Decrement Modeling" with Defense Nuclear Agency/RARP) (U)

Anno, G.H., Wilson, D.B., and Baum, S.J. (1985) Severity levels and symptom complexes for acute radiation sickness: Description and quantification. DNA-TR-86-94. Washington, D.C.: Defense Nuclear Agency, 1985. (U)

Anno, G.H., Wilson, D.B., and Dore, M.A. (1984). Nuclear Weapons Effect Research at PSR, 1983: Acute radiation effects on individual crewmember performance. DNA-TR-85-52. Washington, D.C.: Defense Nuclear Agency, August, 1984. (U)

Bilodeau, E.A. and Bilodeau, I.M. (1969). *Principles of Skill Acquisition*. New York: Academic Press, 1969. (U)

Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Sciences.* New York: Academic Press, 1969. (U)

Dixon, W.J. and Massey, F.J. (1957). *Introduction to Statistical Analysis.* New York: McGraw-Hill, 1957 (2nd. Ed.). (U)

Glickman, A.S. *et. al.* (1983) Estimated effects of intermediate levels of nuclear radiation upon the performance of military tasks: A questionnaire assessment. (Technical Report 82-2/3-1). Norfolk, VA: Organizational Research Group of Tidewater, Inc. [Cited in: Anno, G.H., Wilson, D.B., and Dore, M.A. (1984) Nuclear Weapons Effect Research at PSR, 1983: Acute radiation effects on individual crewmember performance. DNA-TR-85-52. Washington, D.C.: Defense Nuclear Agency, August, 1984.] (U)

Guilford, J.P. (1956). *Fundamental Statistics in Psychology and Education.* New York: McGraw-Hill, 1956. (U)

Kirk, R.E. (1968). *Experimental Design: Procedures for the Behavioral Sciences.* Belmont, CA: Brooks/Cole, 1968. (U)

Mackworth, J.F. (1970). *Vigilance and Attention.* Baltimore: Penguin, 1970. (U)

McClellan, G.E. (1992). Task time data collection for the M198 howitzer. (PSR Report 2322) Santa Monica, CA: Pacific-Sierra Research Corporation, November, 1992. (Prepared under Contracts DNA001-90-C-0139, "Combined Arms Combat Crew Performance Decrement Modeling," DNA001-90-C-0118, and DNA001-90-C-0164, "DNA Improved Casualty Estimation," with Defense Nuclear Agency/RAEM). (U)

McClellan, G.E., Deverill, A.P., and Matheson, L.N. (1994) The effect of MOPP4 on M198 howitzer crew performance. Volume 1—emplacement and displacement times and rates of fire. DNA-TR-93-173-Volume 1. Washington, D.C.: Defense Nuclear Agency, 1994. (U)

McClellan, G.E., Matheson, L.N., and Deverill, A.P. (1994, in preparation) The effect of MOPP4 on M198 howitzer crew performance. Volume 2—task times for fire missions. Santa Monica, CA: Pacific-Sierra Research Corporation, 1994. (Prepared under Contract DNA001-90-C-0164, "DNA Improved Casualty Estimation," with Defense Nuclear Agency/RAEM) (U)

McClellan, G.E. and Wiker, S.F. (1985). A comparison of symptomatology and performance degradation for motion and radiation sickness. DNA-TR-85-354. Washington, D.C.: Defense Nuclear Agency, May, 1985. (U)

Roth, J.T. (1991) A preliminary taxonomy for predicting the magnitude of stressor effects on human task performance. Butler, PA: Applied Science Associates. Working Paper, September 1991. (Prepared under Contract DNA001-90-C-0139, "Combined Arms Combat Crew Performance Decrement Modeling" with Defense Nuclear Agency/RARP) (U)

Roth, J.T. (1992) Reliability and validity assessment of a taxonomy for predicting relative stressor effects on human task performance. (Technical Report 5060-1) Butler, PA: Applied Science Associates, July, 1992. (Prepared under Contract DNA001-90-C-0139, "Combined Arms Combat Crew Performance Decrement Modeling" with Defense Nuclear Agency/RARP) (U)

Rouse, W. (1991). Mental models in training for complex tactical decision making tasks. Report for Contract No. DAAL03-86-D-0001, prepared for Naval Training Systems Center. Norcross, GA: Search Technology, Inc. (U)

APPENDIX A

SELECTED DEMOGRAPHIC CHARACTERISTICS OF CREWMEMBERS

Format Key for Demographic Data Tabulations

Each demographic data record on the following page is in a 94-column format. The following key describes the information and coding of the demographic information

Col 1-4 Record Identification Field (4 digits) 1st digit: irrelevant; 2nd digit: Not used
3rd & 4th digits: Subject # (10-19 Crew 1; 20-29 Crew 2; 30-39 Crew 3)
Col 6-7 Not used

Basic Service Data

Col 9-13 MOS and Grade, e.g, 811E5 (5 Characters; last character is grade)
Col 15-18 Years in Service (x.xx)
Col 20-23 Years in artillery MOS (x.xx)

Combat Experience Data

Col 25 Ever in combat? Y or N; 0 if no answer (1 character)
Col 27-31 When? Month and Year combat service began (MO/YR) (5 characters)
Col 32-35 Duration of combat service (years) (xx.x)

Col 37 Ever in Artillery combat? Y or N; 0 if no answer (1 character)
Col 39-43 When? Month and Year Artillery combat service began (MO/YR) (5 characters)
Col 44-47 Duration of I/A combat service (years) (xx.x)

Specific Crew Position Experience Data

Col 49-50 Ever Chief of Section? (If yes, coded with year individual began occupying position)
Col 52-54 Duration (Months)

Col 56-57 Ever Gunner? (If yes, coded with year individual began occupying position)
Col 59-61 Duration (Months)

Col 63-64 Ever Assistant Gunner? (If yes, coded with year individual began occupying position)
Col 66-68 Duration (Months)

Col 70-71 Ever First Cannoneer? (If yes, coded with year individual began occupying position)
Col 73-75 Duration (Months)

Leadership Experience Data

Col 77 Code for maximum number of M198 crewmembers an individual has supervised:

-1 if no answer	0 if 0	1 if 2	2 if 5
3 if 10	4 if 20	5 if 50	6 if > 50

MOPP 4 Experience Data

Col 79 Any experience in MOPP? Y or N; 0 if no answer (1 Character)
Col 81-82 How many times (xx)
Col 84 In what kind of unit has MOPP been worn? Artillery-1, Infantry-2, Other-3, Basic Training-4
Col 86-89 What is longest time individual has spent in MOPP 4? (hours) (xx.x)
Col 91-94 What is shortest time individual has spent in MOPP 4? (hours) (xx.x)

Table A-1. Demographic data of SME respondents to questionnaires.

CREW 1

1	10	811E4	3.0	2.5	Y			Y			91	12	91	4	90	8	90	9	4	Y	3	1	4.0	0.5.
1	11	811E4	3.0	2.7	Y	12/90	0.5	Y	12/90	0.5	92	5	90	12					3	Y	2	1	2.5	1.0.
1	12	811E3	1.5	1.2	N			N						1		13		1	-1	Y	1	3	0.3	0.3.
1	13	811E3	2.1	1.5	Y	01/91		Y	01/91								91	7	2	Y	1	1	3.0	3.0.
1	14	811E3	1.5	0.9	N			N											1	Y	1	4	0.1	0.1.
1	15	811E3	2.0	1.5	N			N					92	3	91	5	91	4	-1	N				.
1	16	811E3	2.7	2.1	N			N					91	5	91	3	90	2	2	Y	2	1	1.5	1.5.
1	17	811E3	1.3	0.9	N			N											-1	Y	1	4	0.5	0.5.
1	18	811E3	1.7	1.1	N			N											-1	Y	1	4	1.0	1.0.
1	19	811E3	1.5	0.9	N			N											-1	Y	1	2	2.0	0.1.

CREW 2

1	20	811E5	9.5	9.5	Y	91		Y	91		92	36	87	6	87	6	85	6	6	Y	10	1	5.0	0.5.
1	21	811E4	2.5	2.0	Y	04/91		Y	05/91				91	12	90	4			2	Y	8	1	3.0	0.4.
1	22	811E3	0.9	0.6	N			N							92	4			5	Y	2	1	2.0	0.5.
1	23	811E3	0.9	0.6	N			N					92	3	92	0.5			-1	Y	2	1	3.0	.
1	24	811E2	0.9	0.5	N			N											-1	Y	1	2	2.0	.
1	25	811E3	2.6	2.0	N			N											1	Y	3	1	0.5	0.2.
1	26	811E2	1.0	0.6	N			N						7		7		7	-1	Y	1	1	2.0	0.1.
1	27	811E2	1.1	0.7	N			N											4	Y	2	1	2.0	0.5.
1	28	811E3	1.0	0.6	N			N											3	Y	2	1	2.0	1.0.
1	29	811E3	2.5	2.0	Y	91		Y	91						90	24	90	24	5	Y	1	3	3.0	0.1.

CREW 3

1	30	811E5	12.9	12.7	N			N			87	60	86	12	85	12			3	Y	5	1	3.0	0.3.
1	31	811E4	6.0	6.0	N			N			86	12	83	12	83	12	83	12	3	Y	20	1	8.0	1.0.
1	32	811E3	2.2	1.5	N			N					91	2		4			1	Y	4		1.0	0.1.
1	33	811E2	0.9	0.4	N			N									92		-1	N				.
1	34	811E3	1.0	0.2	N			N									92		-1	N				.
1	35	811E2	0.7	0.3	N			N					92		92		92		-1	Y	1	3	1.0	.
1	36	811E4	3.0	3.0	N			N											2	Y	2	1	1.0	0.2.
1	37	811E2	0.6	0.5	N			N											-1	0				.
1	38	811E2	0.6	0.2	N			N											-1	0				.
1	39	811E4	3.1	2.7	N			N							89	30			2	Y	1	1	2.0	.

APPENDIX B

SUMMARY STATISTICS OF SME-ESTIMATE PERFORMANCE INDICES AT 1, 2, AND 4 HOURS IN MOPP 4

Table B-1. Summary Performance Index Statistics for SME Performance Estimates.

Task	At 1 Hour in MOPP 4			At 2 Hours in MOPP 4			At 4 Hours in MOPP 4		
	Mean	Sx	S	Mean	Sx	S	Mean	Sx	S
Receive and Relay Fire Order	0.93	0.02	0.12	0.87	0.02	0.14	0.82	0.02	0.17
Set Deflection on Sight	0.75	0.03	0.24	0.67	0.03	0.26	0.55	0.04	0.28
Traverse Tube and Level Bubbles	0.74	0.03	0.21	0.66	0.03	0.21	0.53	0.03	0.21
Check Sight and Level Bubbles	0.70	0.03	0.24	0.62	0.03	0.25	0.51	0.04	0.26
Set Elevation on Quadrant	0.76	0.03	0.24	0.66	0.03	0.26	0.54	0.04	0.28
Elevate Tube and Level Bubbles	0.68	0.03	0.20	0.58	0.03	0.20	0.47	0.03	0.21
Load Projectile and Powder	0.73	0.03	0.23	0.61	0.03	0.25	0.48	0.04	0.27
Close Breech & Prime	0.77	0.03	0.21	0.63	0.03	0.23	0.54	0.03	0.25
Open Breech and Inspect	0.83	0.03	0.20	0.73	0.03	0.25	0.63	0.04	0.28

APPENDIX C

COMPARISON PLOTS OF SME-ESTIMATE AND MEASURED PERFORMANCE INDICES

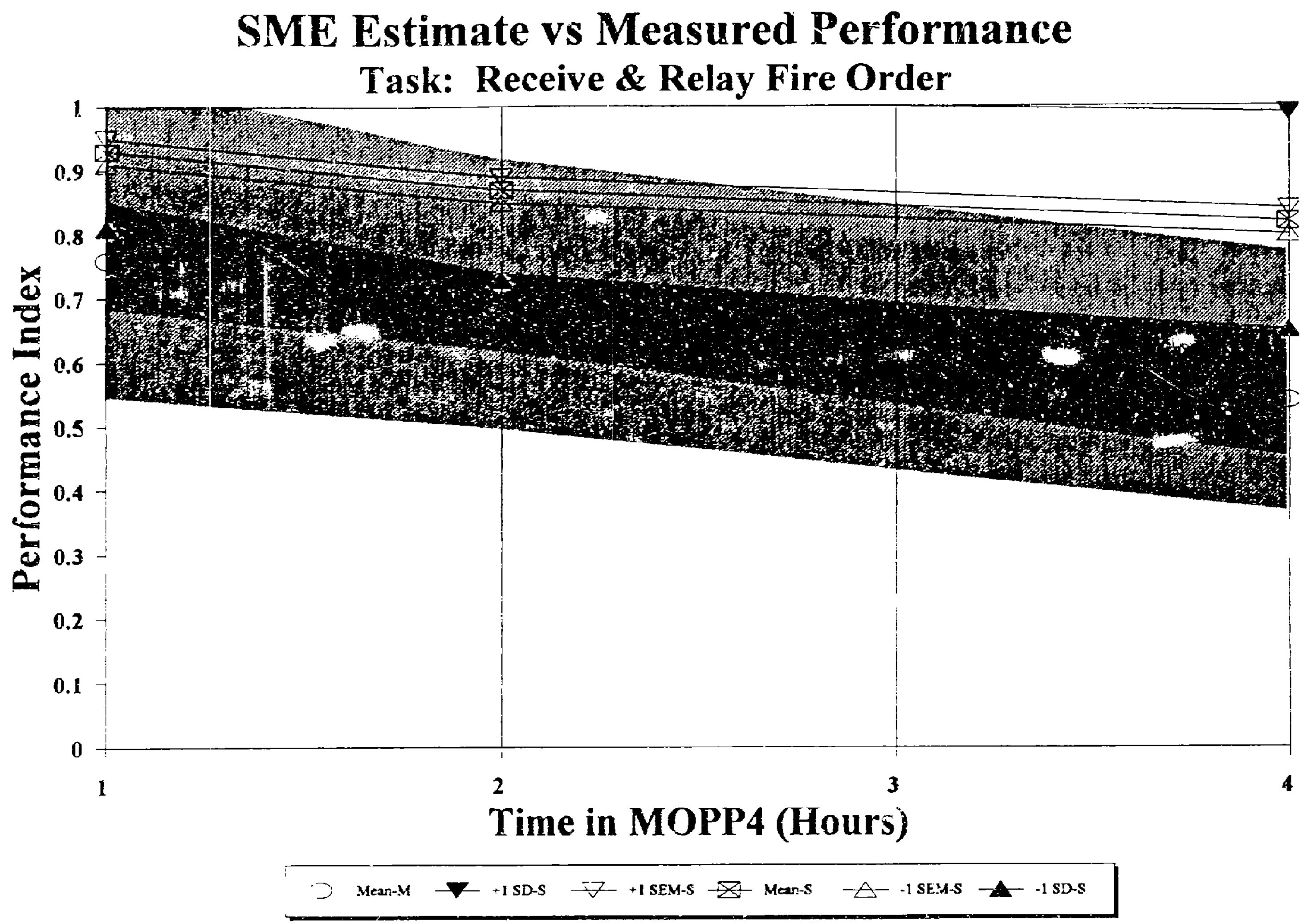

Figure C-1. Comparison plot cf SME-estimate versus measured performance for task "Receive and Relay Fire Order."

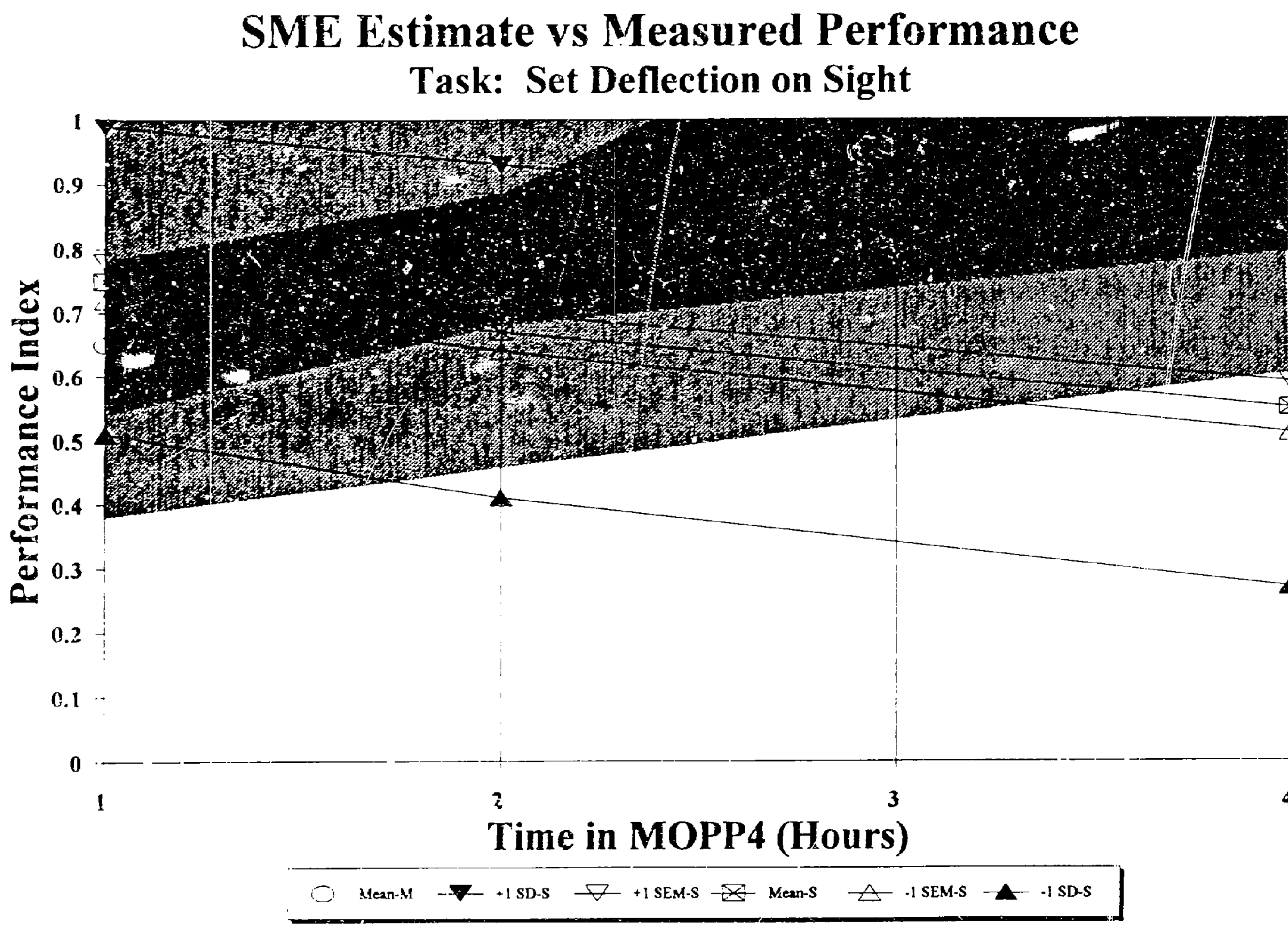

Figure C-2. Comparison plot of SME-estimate versus measured performance for task "Set Deflection on Sight."

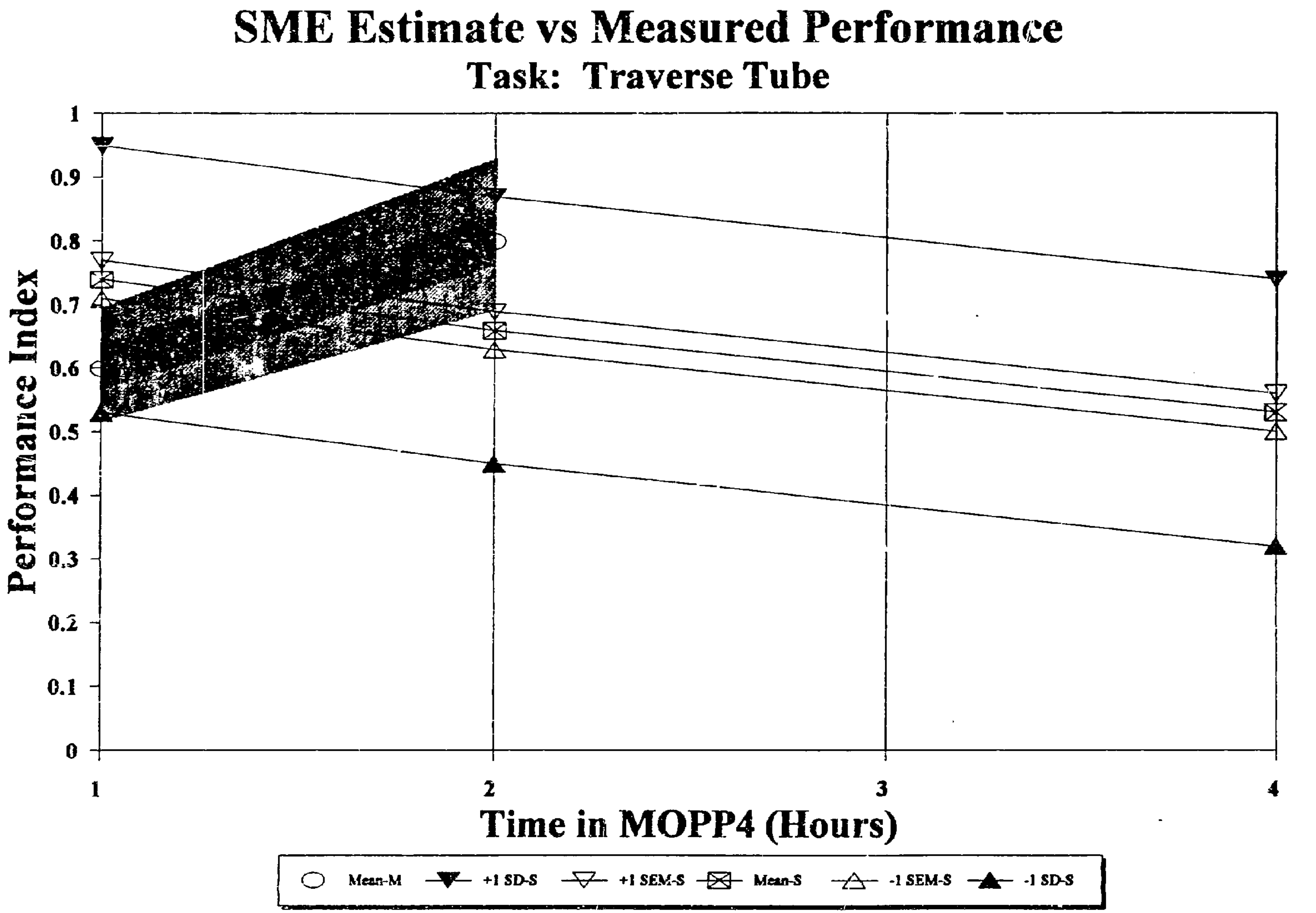

Figure C-3. Comparison plot of SME-estimate versus measured performance for task "Traverse Tube."

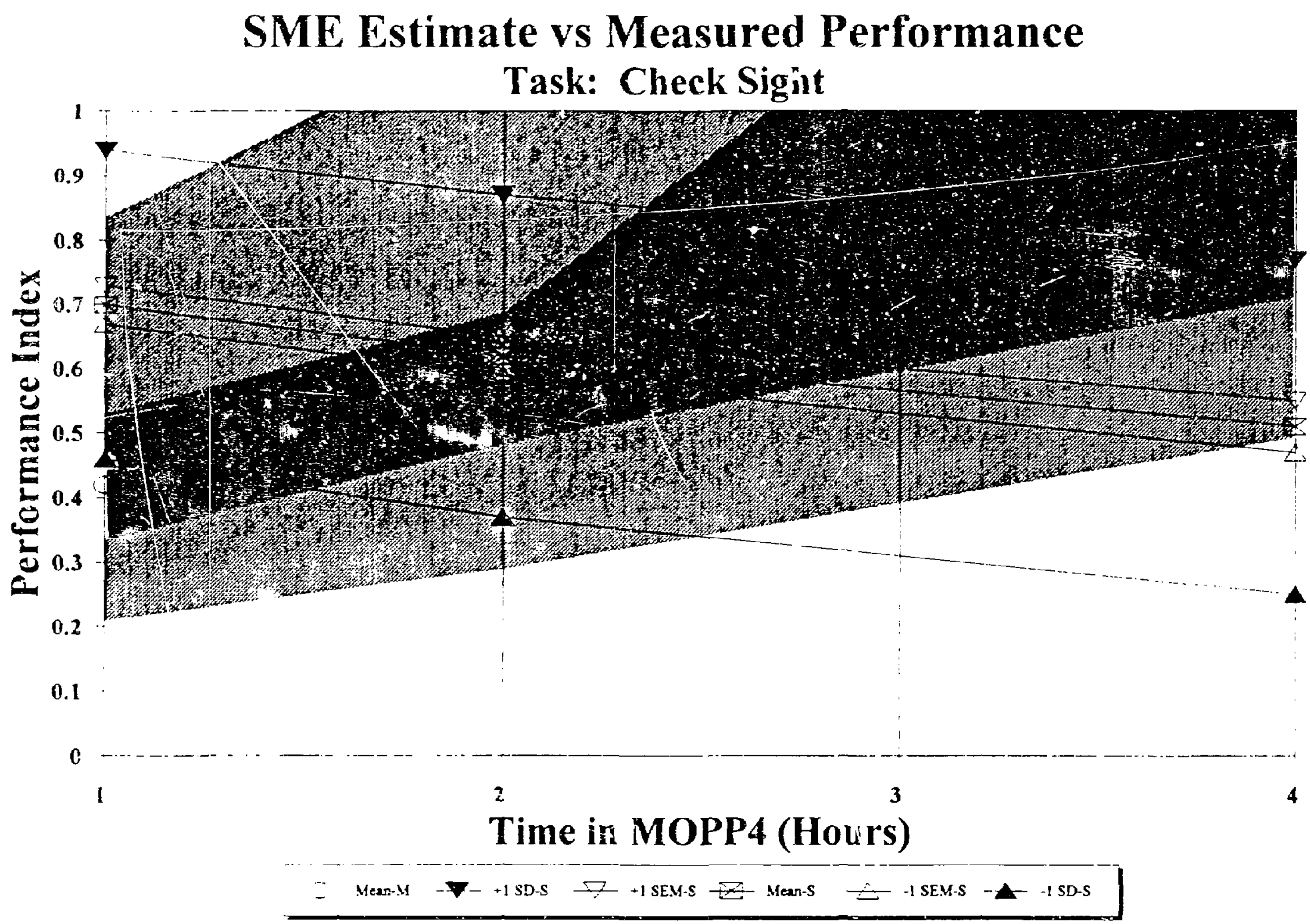

Figure C-4. Comparison plot of SME-estimate versus measured performance for task "Check Sight and Level Bubbles."

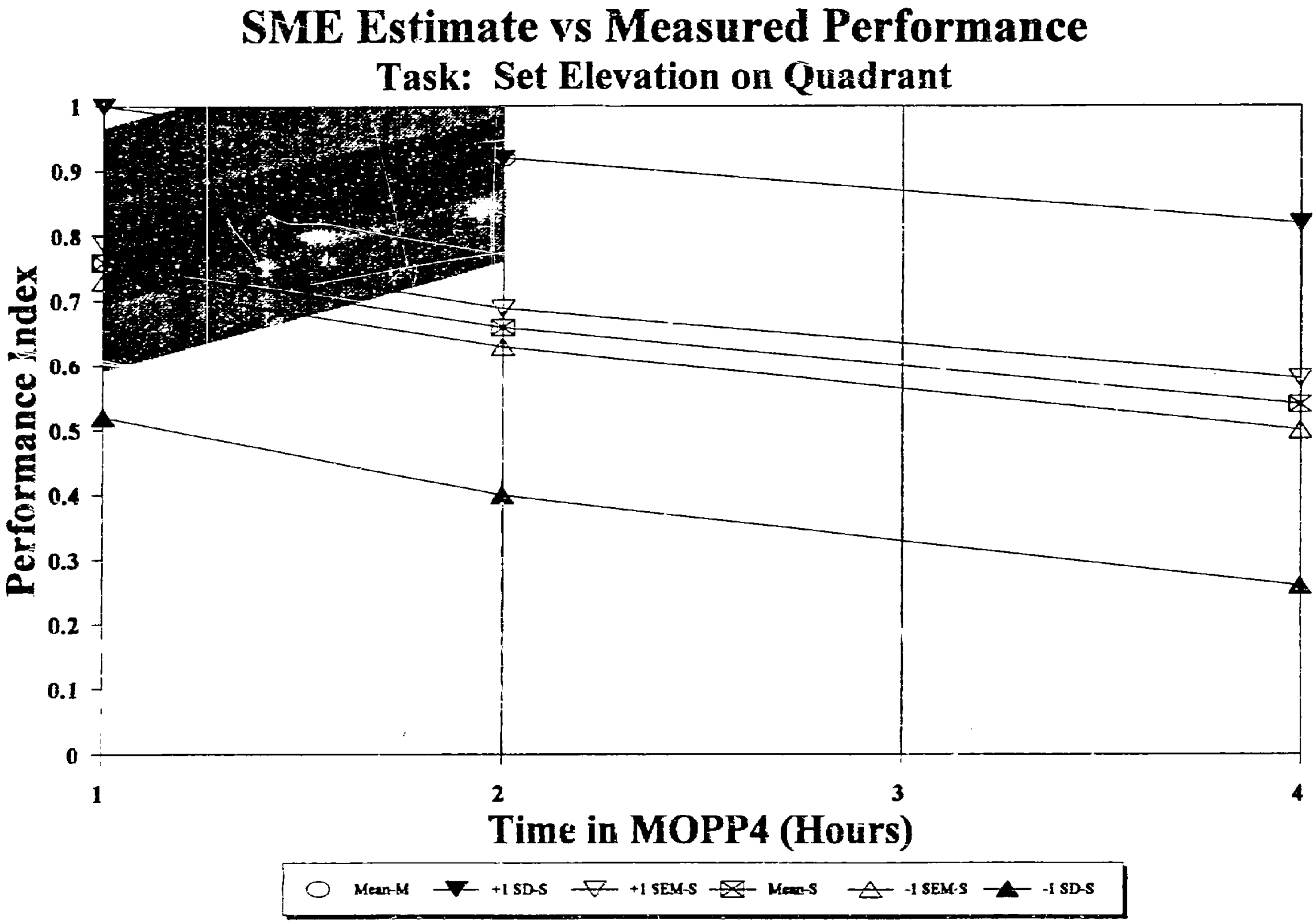

Figure C-5. Comparison plot of SME-estimate versus measured performance for task "Set Evevation on Quadrant."

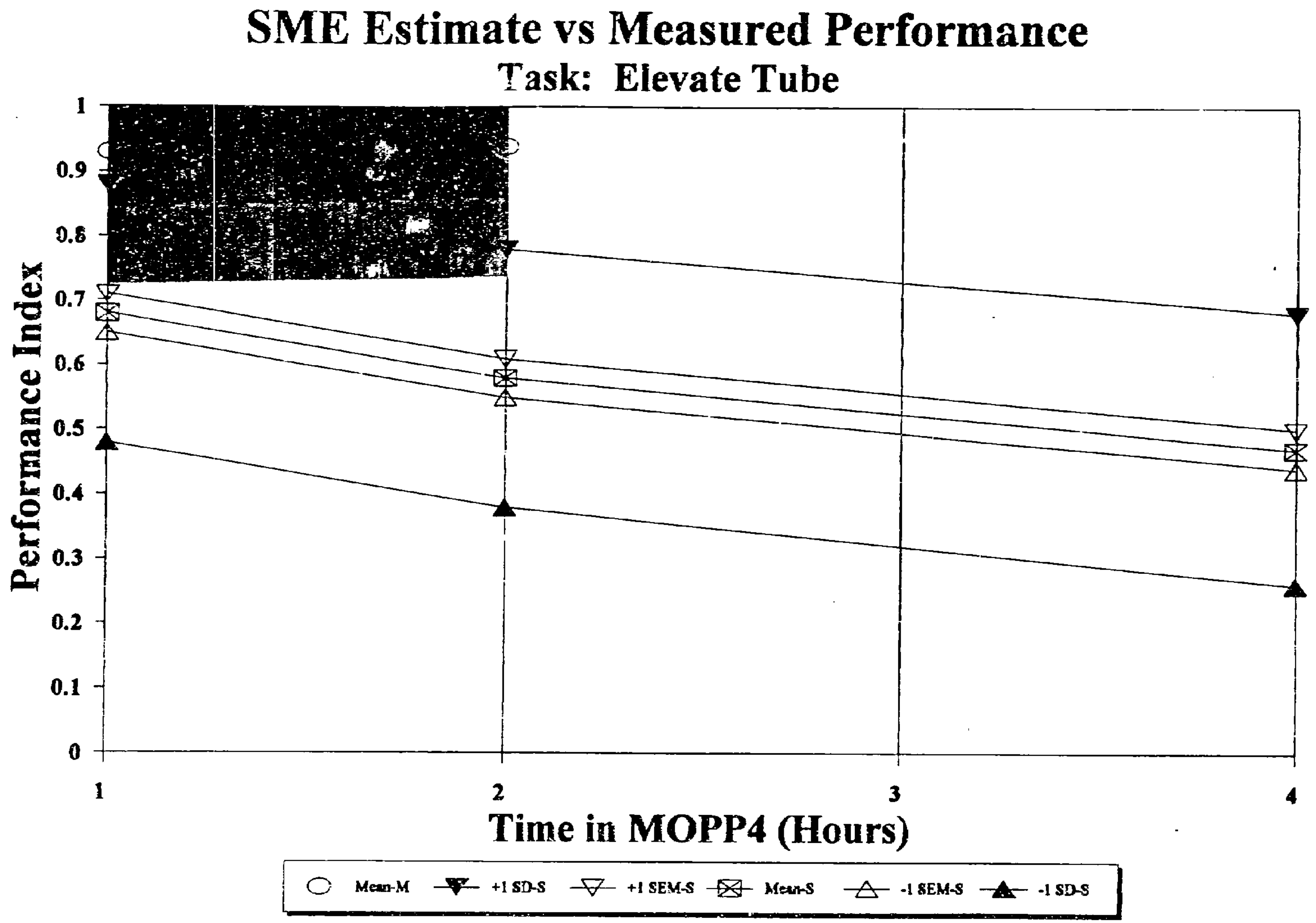

Figure C-6. Comparison plot of SME-estimate versus measured performance for task "Elevate Tube."

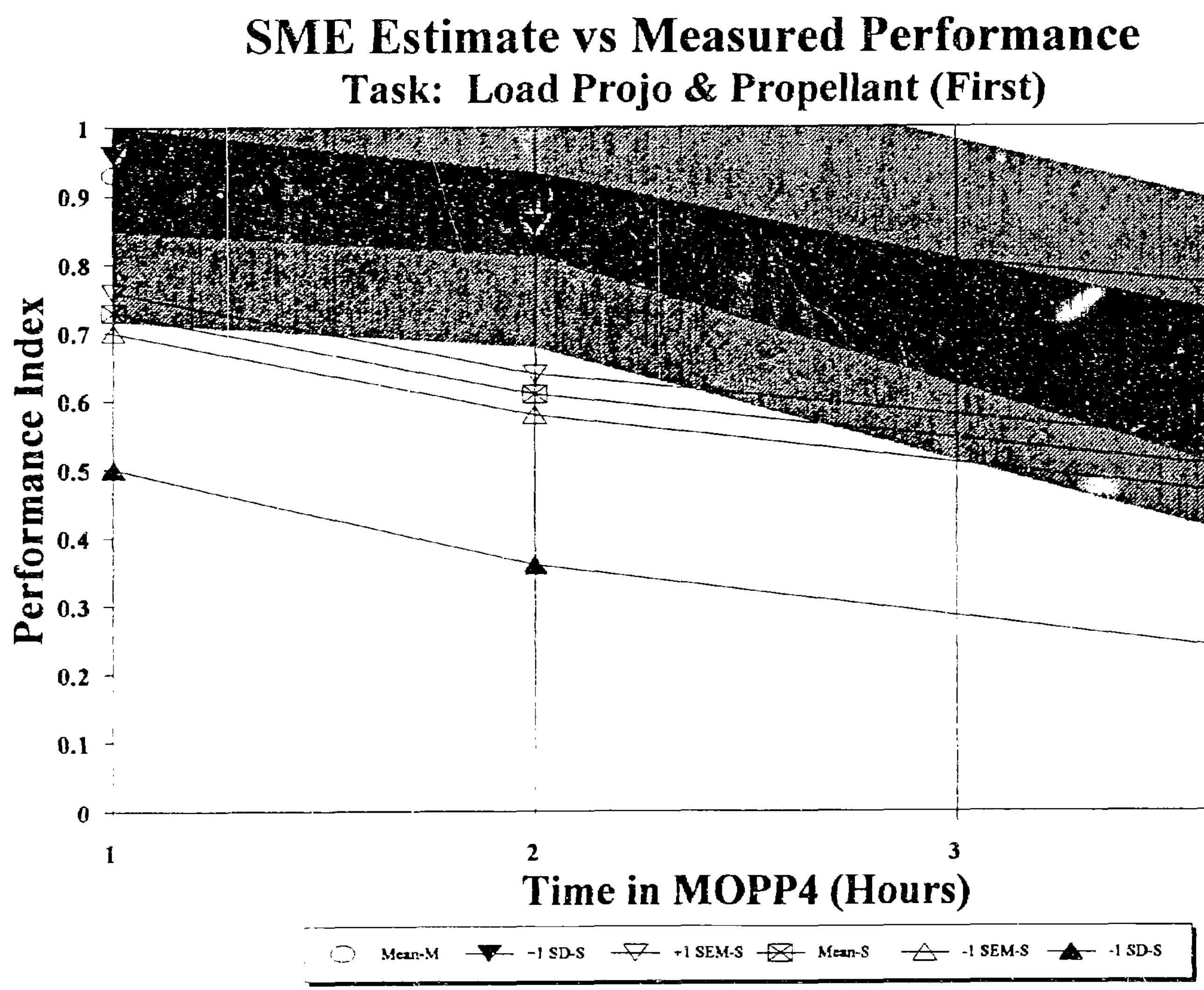

Figure C-7. Comparison plot of SME-estimate versus measured performance for task "Load Projectile and Propellant (First)."

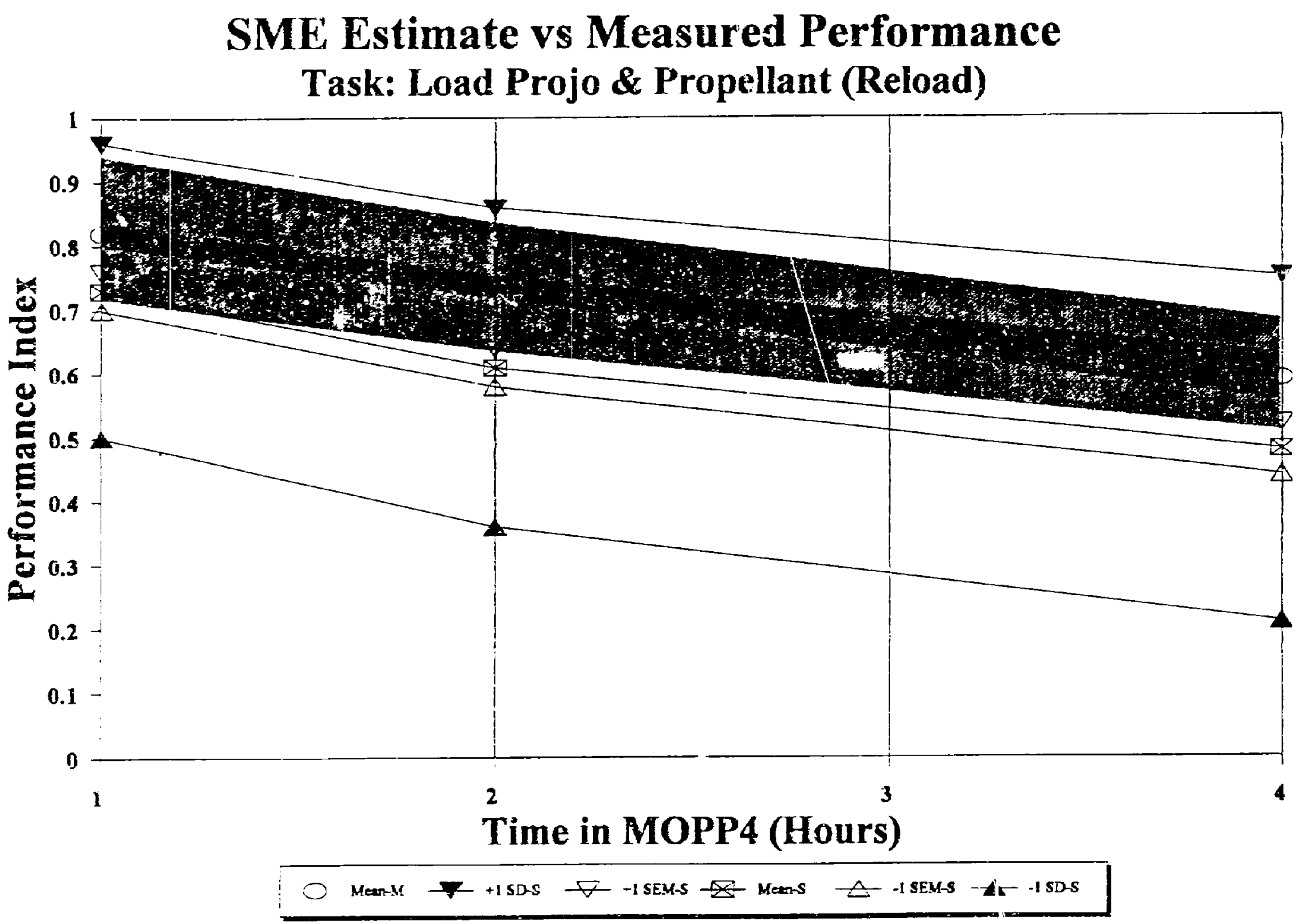

Figure C-6. Comparison plot of SME-estimate versus measured performance for task "Load Projectile and Propellant (Reload)."

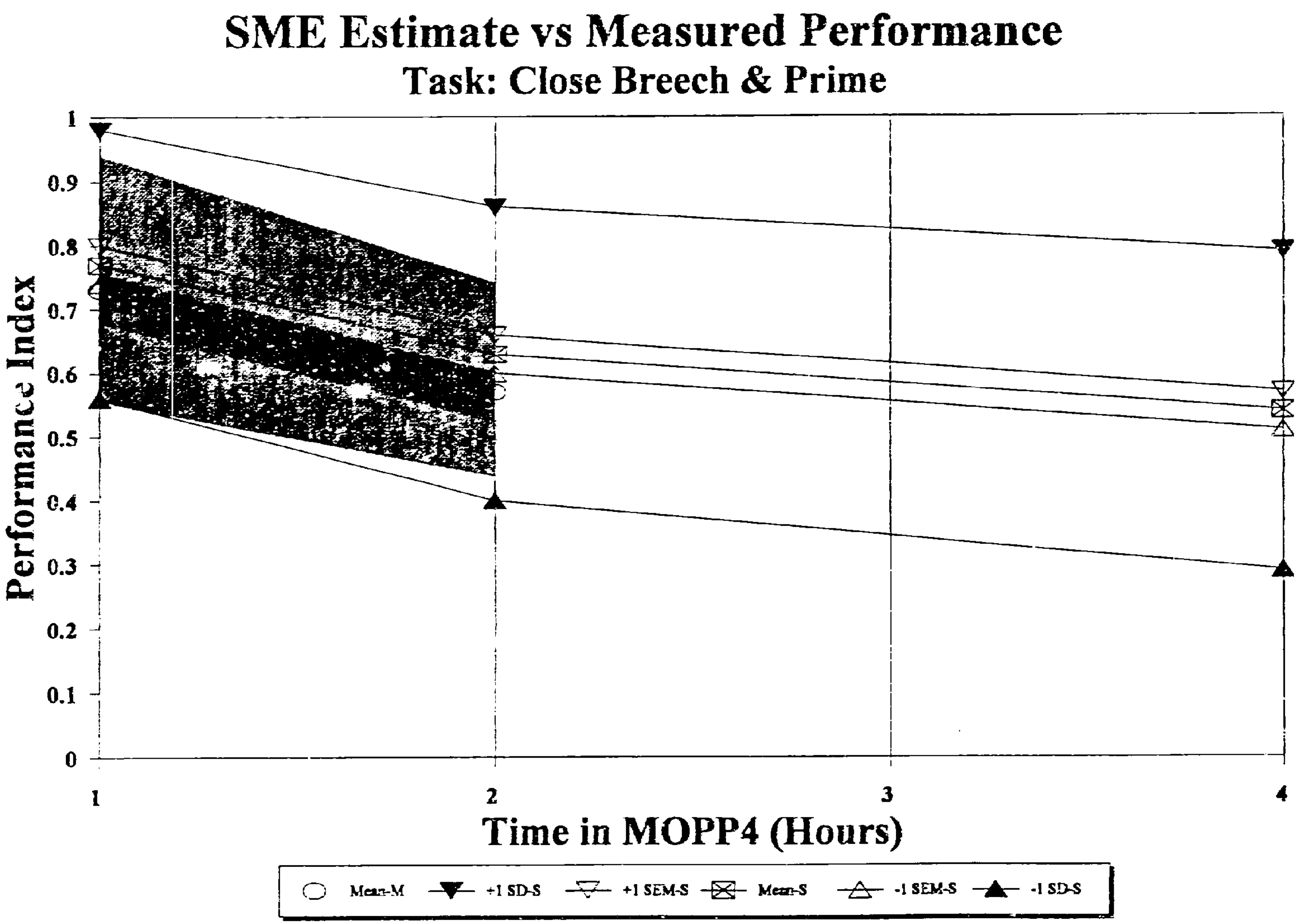

Figure C-9. Comparison plot of SME-estimate versus measured performance for task "Close Breech and P·'me."

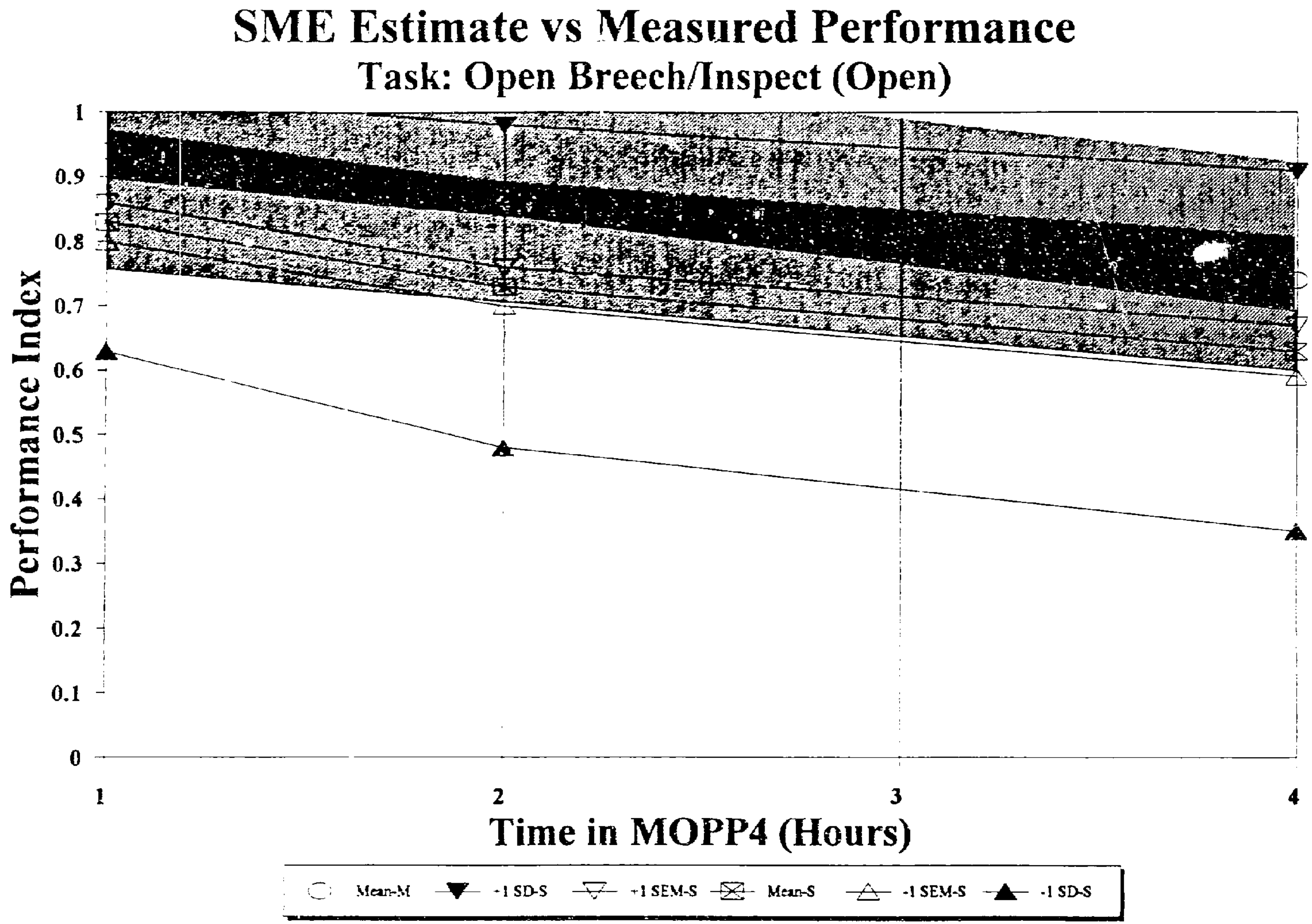

Figure C-10. Comparison plot of SME-estimate versus measured performance for task "Open Breech and Inspect (Open)."

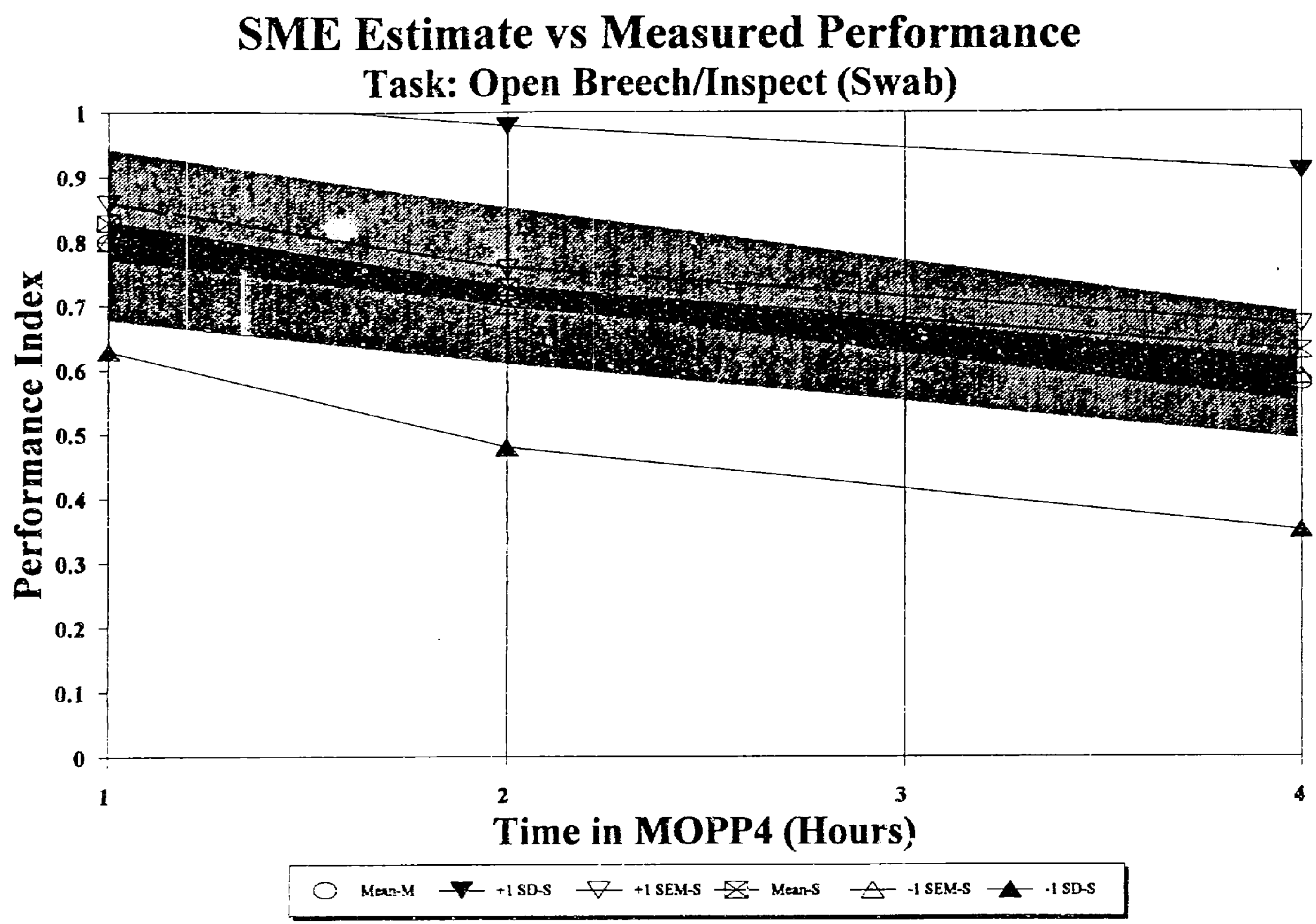

Figure C-11. Comparison plot of SME-estimate versus measured performance for task "Open Breech and Inspect (Swab)."

DISTRIBUTION LIST

DN... -93-173-V3

DEPARTMENT OF DEFENSE

ARMED FORCES RADIOBIOLOGY RSCH INST
ATTN: DIRECTOR
ATTN: MRA
ATTN: SCIENTIFIC DIRECTOR
ATTN: TECHNICAL LIBRARY

DEFENSE INTELLIGENCE AGENCY
ATTN: DB-6B
ATTN: DB-6E

DEFENSE NUCLEAR AGENCY
2 CY ATTN: ISST
ATTN: OPO
ATTN: OPS
ATTN: WE C MCFARLAND
ATTN: WEL
5 CY ATTN: WEP

DEFENSE TECHNICAL INFORMATION CENTER
2 CY ATTN: DTIC/OCP

FIELD COMMAND DEFENSE NUCLEAR AGENCY
ATTN: FCTO
ATTN: TTV
2 CY ATTN: TTV 3416TH TTSQ

NATIONAL DEFENSE UNIVERSITY
ATTN: ICAF TECH LIB
ATTN: NWCLB CR

OASD (C3I)
ATTN: USD/P

TECHNICAL RESOURCES CENTER
ATTN: JNGO

U S EUROPEAN COMMAND
ATTN: ECJ-3
ATTN: ECJ5-N
ATTN: ECJ5N

USSSTRATCOM/J531T
ATTN: J-521

DEPARTMENT OF THE ARMY

ARMY RESEARCH LABORATORIES
ATTN: TECH LIB

NUCLEAR EFFECTS DIVISION
ATTN: STEWS-NE-T

U S ARMY CHEMICAL SCHOOL
ATTN: CARL FRAKER

U S ARMY COMBAT SYSTEMS TEST ACTIVITY
ATTN: JOHN GERDES

U S ARMY COMD & GENERAL STAFF COLLEGE
ATTN: ATZL-SWJ-CA
ATTN: ATZL-SWT-A

U S ARMY CONCEPTS ANALYSIS AGENCY
ATTN: TECHNICAL LIBRARY

U S ARMY FIELD ARTILLERY SCHOOL
ATTN: ATSF-CD

U S ARMY NATIONAL GROUND INTELLIGENCE CTR
ATTN: C WARD

U S ARMY NUCLEAR & CHEMICAL AGENCY
4 CY ATTN: MONA-NU DR D BASH
ATTN: MONA-NU LTC BLISS

U S ARMY RESEARCH LAB
ATTN: AMSRL-SL-B DR KLOPCIC
ATTN: AMSRL-SL-BS E DAVIS
ATTN: SLCBR-D
ATTN: SLCBR-DD-T
ATTN: SLCBR-TB

U S ARMY RESEARCH LABORATORY
ATTN: DIRECTOR
ATTN: DR D HODGE

U S ARMY TEST & EVALUATION COMMAND
ATTN: STECS-NE

U S ARMY VULNERABILITY/LETHALITY
ATTN: AMSLC-VL-NE DR J FEENEY

U S ARMY WAR COLLEGE
ATTN: LIBRARY

U S MILITARY ACADEMY
ATTN: BEHAVORIAL SCI & LEADERSHIP

U S ARMY MEDICAL RESEARCH & DEVELOPMENT COMMAND
ATTN: SGRD-PLE

US ARMY MODEL IMPROVEMENT STUDY MANAGEMENT AGENCY
ATTN: E VISCO

DEPARTMENT OF THE NAVY

MARINE CORPS
ATTN: CODE PPO
ATTN: PSI G RASP

NAVAL WAR COLLEGE
ATTN: DOCUMENT CONTROL
ATTN: LIBRARY

NUCLEAR WEAPONS TNG GROUP, ATLANTIC
ATTN: CODE 222

OFFICE OF CHIEF NAVAL OPERATIONS
ATTN: N 51
ATTN: NIS-22
ATTN: NOP 603
ATTN: NOP 91
ATTN: N455

OPERATIONAL TEST & EVALUATION FORCE
ATTN: COMMANDER

PLANS, POLICY & OPERATIONS
ATTN: CODE-P
ATTN: CODE-POC-30

DEPARTMENT OF THE AIR FORCE

AIR UNIVERSITY LIBRARY
ATTN: LIBRARY

ARMSTRONG LABORATORY
ATTN: AL/CFBE DR F S KNOX

ASSISTANT CHIEF OF STAFF
ATTN: AFSAA/SAMI

DEPUTY CHIEF OF STAFF FOR PLANS & OPERS
ATTN: AFXOOSS

HQ ACC/XP-JSG
ATTN: ACC/XP-JSG

HQ USAFA/DFSELD
ATTN: LIBRARY

DEPARTMENT OF ENERGY

DEPARTMENT OF ENERGY
ATTN: DR T JONES

LAWRENCE LIVERMORE NATIONAL LAB
ATTN: A WARSHAWSKY
ATTN: Z DIVISION LIBRARY

OTHER GOVERNMENT

CENTRAL INTELLIGENCE AGENCY
ATTN: NIO-T
ATTN: N10 STRATEGIC SYS

FEDERAL EMERGENCY MANAGEMENT AGENCY
ATTN: NP-CP
ATTN: OFC OF CIVIL DEFENSE

U S NUCLEAR REGULATORY COMMISSION
ATTN: S YANIV

DEPARTMENT OF DEFENSE CONTRACTORS

ARES CORP
ATTN: A DEVERILL
ATTN: A J BARKLEY
ATTN: G E MCCLELLAN
ATTN: G H ANNO
ATTN: J T ROTH
ATTN: L N MATHESON

HORIZONS TECHNOLOGY INC
ATTN: F GREY

INSTITUTE FOR DEFENSE ANALYSES
ATTN: DOUGLAS SCHULTZ
ATTN: M FINEBERG

JAYCOR
ATTN: CYRUS P KNOWLES

KAMAN SCIENCES CORPORATION
ATTN: DASIAC

LOGICON R & D ASSOCIATES
ATTN: DOCUMENT CONTROL

MICRO ANALYSIS AND DESIGN
ATTN: N LAVINE
ATTN: R LAUGHERY
ATTN: T ROTH

PACIFIC-SIERRA RESEARCH CORP
2 CY ATTN: G ANNO
ATTN: H BRODE

PACIFIC-SIERRA RESEARCH CORP
ATTN: D GORMLEY
2 CY ATTN: G MCCLELLAN

SCIENCE APPLICATIONS INTL CORP
ATTN: D KAUL
ATTN: E SWICK
ATTN: W WOOLSON

SCIENCE APPLICATIONS INTL CORP
ATTN: J MCGAHAN